

Anticipating Groundwater Depletion for Water Scarcity Alleviation through Advanced Deep Learning Models

Abrar Ahmed K¹, Inamul Hussain R Z², Md Sajid³, Rahul B⁴

¹Assistant Professor &HOD /CSE C.Abdul Hakeem College of Engineering and Technology, Melvisharam, Tamil Nadu 632 509, India

²Assistant Professor/CSE C.Abdul Hakeem College of Engineering and Technology, Melvisharam, Tamil Nadu 632 509,India

³Assistant Professor/ECE C.Abdul Hakeem College of Engineering and Technology, Melvisharam, Tamil Nadu 632 509, India

⁴Student/CSE, C.Abdul Hakeem College of Engineering and Technology, Melvisharam, Tamil Nadu 632 509, India

E-mail: ⁴rahul.balaji13289@gmail.com

Abstract

Amidst the challenges posed by climate change, population growth, and pollution, global water availability is diminishing, resulting in critical water shortages. This research proposes a water availability analysis system to tackle the escalating issue exacerbated by urbanization and unsustainable water practices. Conventional methods prove inefficient, prompting the adoption of differentiated analysis through remote sensing and high-resolution mapping for real-time monitoring. The study aims to make a substantial contribution to global water security by providing precise solutions for identifying, managing, and distributing water resources effectively.

Keywords: Department of Science and Technology (DST), Ground Water (GW), Artificial Neural Network, Machine Learning

1. Introduction

India is a country with many famous rivers flowing across the nation and it boasts the world's highest rainfall zone [1]. However, due to unplanned water usage and wastage, several

states in India face severe water shortages during the summer months [3]. As the population continues to rise, the demand for water also increases. With the revolution of artificial intelligence and its advancements in various fields, this study gives insights on water availability in different states in India. According to the statement given by United Nation “A huge number of people may be affected by a water shortage by 2050” [5]. The climate changes also contribute to the change in water precipitation. The water bodies such as rivers, lake, and water spring are related to each other and give impact on the overall availability [6]. The study also focuses on the collecting and managing available groundwater [17]. Though the water was detected and available for use, their availability should prevail on all circumstances hereby we can enforce this by properly managing and distributing the water resources [25].

The study can be facilitated by water estimation that was made to collect and distribute the resources [13]. The detection and future water availability analysis was carried out applying the artificial intelligence-based predictions [10]. The proposed study aims to compare the machine learning models for identifying water using the pre-collected dataset in order to improve the water scarcity problem through proper water management. The rest of the paper is organized with the literature survey in Section 2, the proposed methodology in Section 3, the experimental setup in Section 4, the comparative analysis in Section 5, and the conclusion in Section 6.

2. Literature Survey

Pathak et al. (2014) [24] proposed that the warmer climate may lead to intensification of water cycle and raises evaporation and liquid precipitation. The demand and timing of irrigation and dryness reduces soil moisture content at critical times of the year. Most of the agricultural area in India requires more water and around 2025 global net requirements may increase due to climate change by 3.5-5% by 2025, and 6-6% by 2075. This clearly states that the creation of micro-storage facilities is necessary to improve irrigation as well as recharge the groundwater aquifers.

Riddhi Singh et al. (2019) [25] states the shift from a 1.5°C warmer world to a 2.0°C warmer world would result in a decrease in PCWA for the majority of India's regions. The authors also conduct a sensitivity analysis to determine the relative significance of meteorological and demographic factors in influencing PCWA. They discovered that,

depending on the region's historic backdrop, both population and average annual precipitation are dominant determinants. Regions with historically smaller populations and lower aridity indices are more responsive to population shifts, while regions with larger historical populations are more vulnerable to climatic change. Shi, X et al. (2019) [21] This study presents precipitation nowcasting as a spatiotemporal sequence forecasting issue. Both the input and prediction objective are spatiotemporal sequences. The author presents the ConvLSTM, which incorporates convolutional structures in both input-to-state and state-to-state transitions, and utilize it to create a trainable model for precipitation nowcasting. Our ConvLSTM network consistently beats FC-LSTM and the operational ROVER method for precipitation nowcasting, capturing spatiotemporal correlations more effectively.

V Moosavi et al. (2013) [12] Artificial neural networks (ANNs) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) have been presented as machine learning models with numerous applications in water resource management. Wavelet transformation is a pre-processing approach that can increase forecasting model accuracy by collecting usable data at various resolution levels. The models compared included ANN, ANFIS, Wavelet-ANN, and Wavelet-ANFIS. The study discovered that the models are more accurate for one and two months ahead predictions than for three and four months ahead predictions. M.M. Rahaman et al. (2019) [15] states that the Gravity Recovery and Climate Experiment (GRACE) data is an important global dataset for understanding the variations of groundwater storage in significant aquifers. However, its limited spatial resolution restricts its applicability to the local scale. This work suggests using a Random Forest (RF) model to reduce GRACE-derived groundwater storage anomalies (GWSAs) from $1^\circ \times 1^\circ$ to $0.25^\circ \times 0.25^\circ$ in the Northern High Plains aquifer. This study demonstrates the RF model's ability to generate finer scale GWSA maps for managing groundwater at both local and regional scales, particularly in areas with few groundwater inspection wells.

2.1 Objectives

- To Develop a machine learning model of water availability in India. This model can be used to predict future water availability under different scenarios, such as climate change and population growth. This information can be used to identify the states that are most at risk of water scarcity in the future.

- To develop policy recommendations for improving water availability in India such that they are tailored to the specific needs of different states.
- The statistical and geospatial methods are used to analyse the data and identify patterns and trends in water availability. In order to comprehend the factors that are driving water availability differences across India.

3. Proposed Methodology

Presence of water is the major challenge in country like India. The growth in population exacerbates water scarcity. Water availability can be affected by few factors, including rainfall, evaporation, transpiration, infiltration, runoff, groundwater recharge, water use, and climate change.

3.1 Dataset Description

The historical water availability data were collected from Data portal India website (<https://api.openml.org/d/43813>). This dataset comprises of 689 entries for district wise water availability in India and 16 columns providing various statistical data. The data pertain to the year 2017. In current times, India is going through severe water crisis in many of the regions during the summer [20]. Currently, India is facing a severe water crisis in many regions during the summer [20]. To address this issue, strategies and methods need to be applied for conserving water. Analyzing the dataset can contribute to more efficient distribution of water and regulation of groundwater extraction [19].

3.2 Proposed Architecture

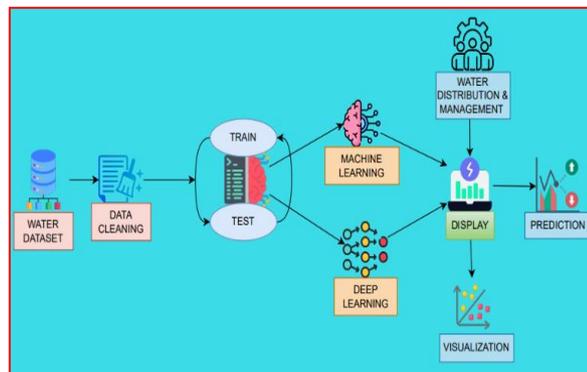


Figure 1. Water Availability Detection

The first step in the machine learning and deep learning process is to collect a large dataset. The dataset collected are cleaned and prepared in the next stage. This may involve removing noise from the data, scaling the data to a consistent range, and converting the data to a format that is compatible with the deep learning framework.

The deep learning model training will be initiated. This involves feeding the data to the model and allowing the model to learn the patterns in the data. The model will learn to adjust its weights and biases in order to minimize the error between its predicted output and the actual output. Once the model is trained, its performance is evaluated on a held-out test set. Evaluating the model on the test set will give an idea of how well the model will generalize to new data. The model from the test set is deployed for predictions on new data. The model predicts the top 3 districts with the highest annual groundwater allocation for domestic use in 2025 they are Cuddalore, Thiruvallur, and Vellore.

4. Experimental Setup

The experimentation process begins with the collection of historical data on water availability, precipitation, runoff, evapotranspiration, groundwater levels, and climate change for the region of interest. This data is then cleaned and prepared, by removing the unwanted data, scaling the features to standard range by applying the Z-score as well as the removing the columns that do not contribute to meaning full information. The dataset was further split for training (70%) and testing (30%). Various machine learning algorithms are then applied to make water availability predictions, with models such as k-nearest neighbours, support vector machines, and artificial neural networks yielding comparable accuracy.

4.1 Support Vector Machine

Support vector machine is a supervised machine learning algorithm which can be implemented for classification and regression techniques. According to the study, SVM algorithm has greater impact on the water availability detection. The algorithm was applied with respect to the feature enlisted for water availability such as precipitation, temperature and soil moisture. As the study focuses on water availability prediction, it can be categorized into two classes like high water availability and low water availability. SVM makes use of hyper

plane for maximizing into two classes. The SVM decision boundary forms a hyperplane that can be separated into two classes of data. Decision limit equation can be written as,

$$w_Tx + b = 0 \tag{1}$$

Where:

W denotes the weight vector

x denotes a data point

b denotes the bias term

The decision boundary equation is used

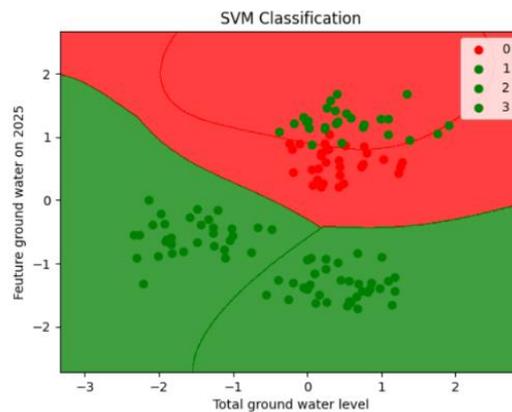


Figure 2. SVM Prediction Plot

The above plot shows the relationship between the total ground water level and the future ground water on 2025. The data was collected from India water availability. The x-axis represents total ground water level and the y-axis represents the future ground water on 2025. The plot shows a positive correlation between the future ground water on 2025 and the total ground water level ($r = 0.75$, $p < 0.01$). This means that level of ground water in the area increases; the total ground water level also increases.

One possible explanation for this correlation is that swimming pools can lower the ground water level by absorbing water from the ground. The water from various districts are typically collected and filled from a municipal water supply, which is sourced from ground water or surface water. As the number of people usage in a district increase, more water is withdrawn from the ground water aquifer.

This can lead to a decrease in the ground water level. Another possible explanation for the correlation is that water collected can raise the ground water level by leaking water into the ground. More research is needed to determine the exact nature of the relationship between the number of people usage in an area and the total ground water level of a state. However, the scatter plot provides some evidence that there is a positive correlation between these two variables.

4.2 Random Forest Classifier

A random forest classifier algorithm works by building a set of decision trees, each tree was trained on random data subset. Finally, the trees were combined to produce final prediction. Random forest classifiers were known for their accuracy and robustness. They are easy to interpret and understand the factors that impacts on water availability. The random forest was utilized in the prediction for their features like accuracy even with the noisy data, robustness in fitting with any data comparative to other algorithms.

A study in India supports that random forest classifier produces good results in water availability prediction in a reservoir. The prediction made by the random forest classifier does not get affected on even any sudden unexpected change in climate. This makes random forest classifier as the high accuracy producing algorithm.

Classification Report

Accuracy: 0.944

Table 1. Classification Report of Random Forest Classifier

CL Report:	Precision	Recall	f1-score	Support
0	0.90	0.87	0.89	31
1	0.86	0.89	0.87	27
2	1.00	1.00	1.00	32
3	1.00	1.00	1.00	35
Accuracy	-	-	0.94	125
Macro Avg.	0.94	0.94	0.94	125
Weighted Avg.	0.94	0.94	0.94	125

The random forest classifier is the best algorithm for making analysis with the large scale of data. This is the reason; the algorithm facilitates with the good results.

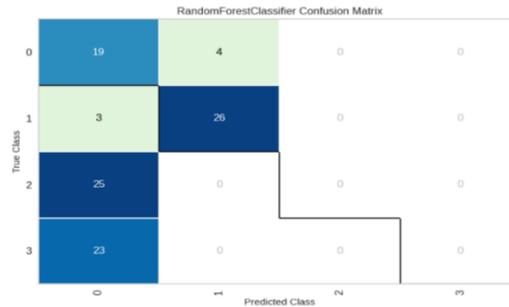


Figure 3. Confusion Matrix Plot

The matrix shows how well the model performed on a given dataset, with the rows representing the actual classes of the data and the columns representing the predicted classes. The diagonal elements of the matrix represent the number of correct predictions, while the off-diagonal elements represent the number of incorrect predictions.

The diagonal elements of the matrix are all relatively high, indicating that the model is making a lot of correct predictions. The off-diagonal elements of the matrix are all relatively low, indicating that the model is making a relatively small number of incorrect predictions. It is making a lot of correct predictions and a relatively small number of incorrect predictions. This is indicated by the high diagonal elements and low off-diagonal elements of the matrix. The random forest classifier algorithm mentioned yields 92% accuracy, which is also a good indication that the model is performing well. This means that it is correctly predicting the class of 92% of the data points in the dataset. Overall, the confusion matrix and accuracy score suggest that the Random Forest classifier is a good model for the task at hand. It is making accurate predictions and is likely to generalize well to new data.

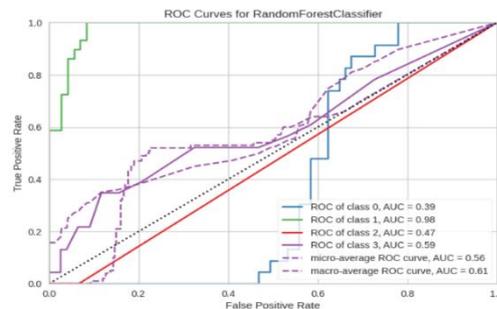


Figure 4. ROC Curves

An ROC curve, or receiver operating characteristic curve, is a graph that shows the performance of a classification model at all classification thresholds. The ROC curve plots two parameters:

- True Positive Rate (TPR): This is the percentage of positive cases that are correctly identified by the model.
- False Positive Rate (FPR): This is the percentage of negative cases that are incorrectly identified by the model as positive.

The TPR and FPR are calculated for a range of classification thresholds. The classification threshold is the value that the model's output must exceed in order for the model to predict that the data point that belongs to the positive class. The ROC curve is a useful tool for evaluating the performance of classification models because it shows how the TPR and FPR change as the classification threshold is varied. This allows us to find a classification threshold that gives us the best balance between TPR and FPR. The performance of an ROC curve. The AUC ranges from 0 to 1, and a higher AUC indicates better performance. The AUC for the Random Forest classifier is 0.98, which is very good. Overall, the ROC curve shows that the Random Forest classifier is a well-performing classification model. It has a high TPR and a low FPR, and it has a high AUC. This indicates that the classifier is a well-performing classification model. The plot show shows good true positive rate (TPR) and false positive rate (FPR) results and it is supportive for the algorithm. The plot states the ROC curve values in range of 0 to 1 which ensemble that this is the perfect classifier.

4.3 Artificial Neural Network (ANN)

Artificial neural networks (ANNs) are a type of machine learning algorithm that can be used to predict water availability. ANNs are trained on historical data on water availability and other relevant factors, such as climate, land use, and population. Once trained, ANNs can be used to predict water availability for future time periods in a variety of settings, including river basins, reservoirs, and agricultural regions. For example, ANNs have been used to predict water availability during droughts, which can help water managers to allocate water resources more effectively.

5. Comparative Analysis

The comparison of accuracy was tabulated from the applied four machine learning models with respective to data taken for analysis.

Table 2. Accuracy Comparison Tabulation

Data size	KNN algorithm	SVM algorithm	Random Forest algorithm	Artificial Neural Network
690	78%	93%	95%	85%

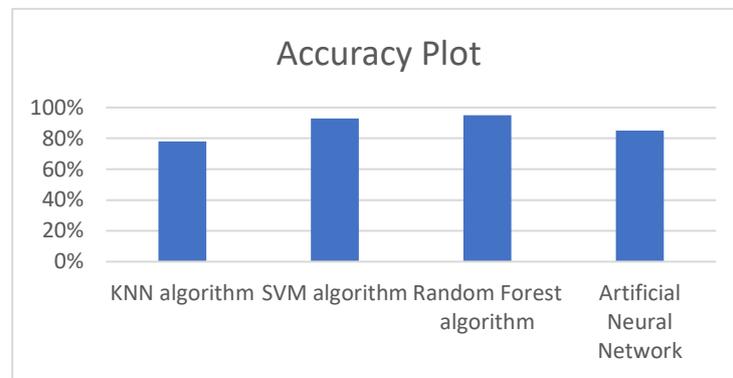


Figure 5. Comparison of Accuracy

The bar plot shows the accuracy of different algorithms on a classification task. The x-axis of the bar graph shows the algorithm name, and the y-axis shows the accuracy of the algorithm. The bar graph shows that the Random Forest algorithm has the highest accuracy, followed by the Support Vector Machine (SVM) algorithm, Artificial Neural Network (ANN) and K-Nearest Neighbors (KNN) algorithm.

The bar graph also shows that the accuracy of all of the algorithms is relatively high, with all of the algorithms achieving an accuracy of at least 78%. This suggests that all of the algorithms are well-suited to the classification task. The performance of four different classification algorithms was evaluated on a dataset of District-wise-Ground-Water-Resources-by-July-2017. The Random Forest algorithm achieved the highest accuracy (95%), followed by the Support Vector Machine (SVM) algorithm (93%), and the ANN algorithm (85%). The KNN achieved the lowest accuracy (78%).

The comparative analysis bar chart represents accuracy comparison with four algorithms. The accuracy continues to vary when large numbers of dataset were taken. As we can state that number of entries directly affects the accuracy of the algorithm.

5.1 Prediction Plot Shows Ground Water Level in 2050

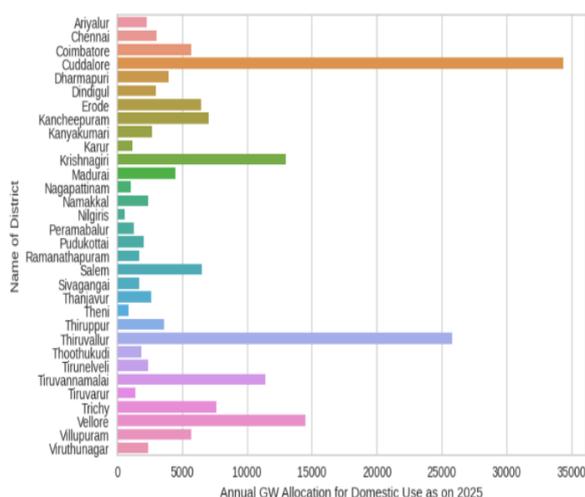


Figure 6. 2025 AnnualGW for Domestic use Prediction Plot

The plot shows the annual GW allocation for domestic use as of 2025 in different districts in Tamil Nadu. The x-axis of the plot shows the district name, and the y-axis shows the GW allocation in cubic meters per capital. The plot shows that there is a large variation in the GW allocation for domestic use across different districts in Tamil Nadu. Some districts, such as Chennai and Coimbatore, have a relatively low GW allocation, while other districts, such as Villupuram, Thiruvallur and Cuddalore, have a relatively high GW allocation. The plot also shows that there is no clear pattern to the variation in GW allocation. For example, districts that are located in the same region can have very different GWP allocations. For example, the Cuddalore district has a GW allocation of 150 cubic meters per capital, while the Chennai district, has a GW allocation of only 50 cubic meters per capital.

The Government of India allocates ground water potential (GWP) for domestic use to different districts based on a number of factors, including the population of the district, the level of economic development in the district, the availability of water resources in the district, the climate of the district, and the government policies regarding water allocation. There is a need for more research to understand the factors that are affecting the GWP allocation in each

district and to develop policies to ensure that all districts have equitable access to water resources.

There are a number of factors that can affect the GW allocation for domestic use in a given district. These factors include:

- The population of the district
- The stage of economic development in the district
- The availability of water resources in the district
- The climate of the district
- The government policies regarding water allocation

As a result, the analysis data interprets that Cuddalore and Thiruvallur districts in Tamil Nadu may meet best annual ground water rise in the year of 2025. The interpretation has made with the data collected on the 2017 ground water analysis.

6. Conclusion

This proposal has examined water availability in various states in India. The study found that water availability is unevenly distributed across the country, with some states experiencing severe water scarcity. The study also identified a number of factors that influence water availability, including climate, topography, population, and water management practices. The experimentation concluded that districts with low groundwater (GW) allocation, such as Chennai and Coimbatore, will need to focus on rainwater harvesting and groundwater recharge, reduce water consumption, improve water efficiency in agriculture and industry, and consider desalination to supplement water supplies. Water distribution in these districts should prioritize drinking water and essential needs, ration water to non-essential users, and invest in water distribution infrastructure to reduce losses. Districts with high GW allocation, such as Villupuram, Thiruvallur, and Cuddalore, will need to use groundwater sustainably, avoid overexploitation, protect groundwater quality from pollution, improve water efficiency in agriculture and industry, and consider artificial recharge of groundwater. Water distribution in

these districts should prioritize drinking water and essential needs, ensure equitable distribution of water to all users, and invest in water distribution infrastructure to reduce losses.

6.1 Future work

This research paper has provided a valuable overview of water availability in various states and districts in India. However, there is still much scope for future research in this area. The further implementation can be done to develop more refined water availability models that can predict future water availability under different scenarios, such as climate change and population growth, Interpret on more case studies of specific states or regions to examine the specific factors that influence water availability and the challenges and opportunities for improving water management, and in conclusion to this, The development and evaluation can be established for water availability and management policies and interventions for India.

References

- [1] Lo, M.H.; Famiglietti, J.S.; Yeh, P.F.; Syed, T. Improving parameter water table depth simulation and parameter estimation in land surface water storage. *Water Resource. Res.* 2010, 46.
- [2] Zhou, T.; Wang, F.; Yang, Z. Comparative analysis of ANN and SVM models combined with wavelet pre-process for groundwater depth prediction. *Water* 2017, 9, 781.
- [3] Adamowski, J.; Fung Chan, H.; Prasher, S.O.; Ozga-Zielinski, Fung Chan, H.; Prasher, B.; Sliusarieva, A. Comparison of autoregressive integrated moving average, artificial neural network, multiple linear and nonlinear regression, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resource. Res.* 2012, 48.
- [4] Sahoo, S.; Jha, M.K. Groundwater-level prediction using artificial neural network techniques and multiple linear regression: A contrast assessment. *Hydrogeol. J.* 2013, 21, 1865–1887.

- [5] Bourennane, H.; King, D.; Couturier, A. Comparison of kriging with simple and external drift and linear regression to predict soil horizon thickness with different sample density. *Geoderma* 2000, 97, 255–271.
- [6] Tiwari, M.K.; Adamowski, J. Urban water uncertainty assessment and demand forecasting using ensemble wavelet-bootstrap-neural network models. *Water Resource Res.* 2013, 49, 6486–6507.
- [7] Arandia, E.; Ba, A.; Eck, B.; McKenna, S. Tailoring seasonal time series models to forecast short-term water demand. *J. Water Resour. Plan. Manag.* 2016, 142, 04015067.
- [8] Nielsen, A. *Practical Time Series Analysis: Prediction with Statistics and Machine Learning*; O'Reilly: Sebastopol, CA, USA, 2020.
- [9] Yoon, H.; Jun, S.C.; Hyun, Y.; Bae, G.O.; Lee, K.K. A comparative study of support vector machines and artificial neural networks for predicting groundwater levels in a coastal level aquifer. *J. Hydrol.* 2011, 396, 128–138.
- [10] Sun, A.Y. Predicting groundwater level changes using GRACE data. *Water Resour. Res.* 2013, 49, 5900–5912.
- [11] Emamgholizadeh, S.; Moslemi, K.; Karami, G. Prediction the groundwater level of bastam plain (Iran) by adaptive neuro-fuzzy inference system (ANFIS) and artificial neural network (ANN). *Water Resour. Manag.* 2014, 28, 5433–5446.
- [12] Moosavi, V.; Vafakhah, M.; Shirmohammadi, B.; Behnia, N. Hybrid wavelet-ANFIS model for predicting groundwater levels across timeframes *Water Resources. Manag.* 2013, 27, 1301–1321.
- [13] Dos Santos, C.C.; Pereira Filho, A.J. Predicting water needs for São Paulo: A forecasting model for the metropolitan area of Brazil. *Water Resour. Manag.* 2014, 28, 4401–4414.
- [14] Huang, F.; Huang, J.; Jiang, S.H.; Zhou, C. Chaotic groundwater prediction with support vector machines. *J. Hydroinform.* 2017, 19, 586–606.

- [15] Rahaman, M.M.; Maheshwari; Thakur, B; Kalra, A.; Li, R; P. Precision Pumping GRACE: Random Forests for High-Resolution Storage Estimates. *Environments* 2019, 6, 63.
- [16] Jing, W.; Yao, L.; Zhao, X.; Zhang, P.; Liu, Y.; Xia, X.; Song, J.; Yang, J.; Li, Y.; Zhou, C. Understanding the decline of freshwater resources in the Yellow River Basin. *J. Geophys. Res. Atmos.* 2019, 124, 12963–12984.
- [17] Jing, W.; Zhao, X.; Yao, L.; Di, L.; Yang, J.; Li, Y.; Guo, L.; Zhou, C. Can terrestrial water storage dynamics be estimated from climate anomalies? *Earth Space Sci.* 2020, 7, e2019EA000959.
- [18] Sahour, H.; Sultan, M.; Vazifedan, M.; Abdelmohsen, K.; Karki, S.; Yellich, J.A.; Gebremichael, E.; Alshehri, F.; Elbayoumi, T.M. Boosting water resource management: Statistical downscaling and gap-filling of GRACE data for local decision-making. *Remote Sens.* 2020, 12, 533.
- [19] Mukherjee, A.; Ramachandran, P. Predicting Indian groundwater levels using GRACE data: Comparing SVR, ANN and LRM for irregular time series. *J. Hydrol.* 2018, 558, 647–658.
- [20] Seyoum, W.M.; Kwon, D.; Milewski, A.M. Revolutionizing groundwater analysis: Machine learning unlocks the power of GRACE GWSA data for detailed insights into a glacial aquifer system. *Remote Sens.* 2019, 11, 824.
- [21] Shi, X.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Published in the proceedings of the 2015 Advances in Neural Information Processing Systems conference (Montreal, Canada)
- [22] Shi, E.; Li, Q.; Gu, D.; Zhao, Z. Revolutionizing weather forecasting: A groundbreaking method for radar echo extrapolation with convolutional neural networks. In *Multimedia Modelling (MMM 2018)*; Schoeffmann, K., Ed.; Springer: Cham, Switzerland, 2018; Volume 10704, pp. 16–28.

- [23] Shi, X.; Gao, Z.; Lausen, L.; Wang, H.; Yeung, D.Y.; Wong, W.k.; Woo, W.C. Deep learning for precipitation now casting: Advancing the accuracy of rain forecasting: A new deep learning approach and performance benchmark and Neural Information Processing Systems, Long Beach, USA, CA. 4–9 December 2017; pp. 5617–5627.
- [24] Pathak, H., P. Pramanik, M. Khanna, and A. Kumar. "Climate change and water availability in Indian agriculture: impacts and adaptation." *Indian Journal of Agricultural Sciences* 84, no. 6 (2014): 671-679.25. Singh, Riddhi, and Rohini Kumar. "Climate versus demographic controls on water availability across India at 1.5 C, 2.0 C and 3.0 C global warming levels." *Global and planetary change* 177 (2019): 1-9.