Explainable AI for Ground water nitrate content prediction using machine learning Frameworks in Telangana

Keyur Padia
Department of Computer Science
Vellore Institute Of Technology
Vellore, India
keyur.padia2022@vitstudent.ac.in

Anshul Sunkari
Department of Computer Science
Vellore Institute Of Technology
Vellore, India
anshul.sunkari2022@vitstudent.ac.in

Abhinav Gupta
Department of Computer Science
Vellore Institute Of Technology
Vellore, India
abhinav.gupta2022@vitstudent.ac.in

C. L. BIJI*
Department of Computer Science
Vellore Institute Of Technology
Vellore, India
biji.cl@vit.ac.in

***Abstract*—Groundwater contamination is a huge risk to public health and environmental sustainability. The present study focuses on forecasting nitrate levels in groundwater in Telangana, India, using explainable AI (XAI) method. The present work utilizes linear regression, polynomial regression, decision tree regressor, random forest regressor, Support vector machine regressor, Multilayer Perceptron Regressor, XGBoost Regressor, LightGBM, and KNN regressor to forecast nitrate levels according to numerous hydrogeological and environmental factors. The multilayer perceptron regression model was reported to have superior performance compared to all other regressor models with a mean absolute error of 18.879 and a root mean square error of 27.703 To ensure the interpretability and transparency of the model, the SHAP method XAI (SHapley additive causes) has been utilized in multilayer perceptron regression, offering insight into the characteristic importance of the allied features responsible for nitrate contamination. The framework greatly improves prediction dependability and interpretability, allowing policy makers and environmental agencies to make informed decisions for the administration of groundwater.**

I. Introduction

Groundwater is the most important source of freshwater. Around 70% of groundwater withdrawn worldwide is used for agriculture [1]. In India, around 80% of the pastoral population and 50% of the communal population use groundwater for domestic purposes. Groundwater quality is a pressing global concern. In India, there are several problems related to the quality of groundwater water, including fluoride and nitrate, which are all dangerous to mortal health [2]. Balancing nitrate use with sustainable farming, safe pesticide practices, and strong water programs is vital to protect groundwater. [3].

A study conducted in Medchal District, Telangana, revealed high levels of groundwater contamination by nitrates, posing health risks to both children and adults. The overuse of chemical fertilizers by farmers and improper disposal of wastewater from fields have been identified as major contributors to nitrate contamination in the region [4]. The World Health Organization recommends a maximum nitrate level of 50 mg/L in drinking water to prevent health risks such as methemoglobinemia, or 'blue baby syndrome'. In India, the Bureau of Indian Standards has set a slightly stricter standard of 45 mg/L. These guidelines are established to protect public health, especially in areas where nitrate contamination of groundwater is a concern due to agricultural runoff and other sources of pollution. It's important for water quality monitoring to adhere to these standards to ensure the safety of drinking water supplies [6].

Groundwater modeling is a vital tool for organizing hydrologic data, understanding system behavior, and predicting responses to external stresses [7]. Machine learning models have significantly advanced groundwater quality prediction and management [8]. Ensemble algorithms such as Random Forest and XGBoost show strong performance in predicting nitrate concentrations [9]. Various factors have been linked as influential in predicting nitrate attention, including position, nitrogen situations, ammonium, phosphate, pH, temperature, dissolved oxygen, natural oxygen demand, suspended solids, and streamflow [10].

This research formulates and implements a sophisticated, explainable machine learning approach to forecast nitrate levels in Telangana groundwater based on a multi-year record from wells across the state. The method blends high prediction quality with explainability so that informed decision support is possible for researchers and policymakers attempting to deal with nitrate contamination.

II. Materials and methods

A. Case Study

Telangana is a country located in the southern portion of India, covering an area of about 112,077 square kilometers [11]. It shares borders with Maharashtra, Chhattisgarh, Karnataka, and Andhra Pradesh [12]. The state lies on the Deccan Plateau, which is marked by undulating terrain and scattered hills [13]. The main rivers that pass through Telangana are Godavari, Krishna, Bhima, Manjira, and Musi [14]. The Godavari River is the second-longest river in India, which flows into

Telangana from Maharashtra and provides good-sized water resources [15]. The state receives an average annual rainfall of approximately 905 mm, with significant spatial variability due to monsoonal influences [16]. The southwestern monsoon accounts for almost 80% of the total annual rainfall [17]. The groundwater ranges in Telangana vary widely. Overexploitation in some districts has led to depletion, with water tables falling below 20 meters in some areas [18].

However, areas with adequate aquifers, such as Nizamabad and Karimnagar, have exceptionally robust groundwater levels [19]. Telangana has many prominent dams, including Nagarjuna Sagar Dam, Sriram Sagar Dam, and Singur Dam, which are essential for irrigation and water storage. The state also has various aquifers, in particular fractured granite and basaltic formations. The elevation of Telangana varies from 100 to 900 meters above sea level. The highest point is located in the Anantagiri Hills at about 1,168 meters. Telangana has a primarily tropical climate, classified as semi-arid in some areas and sub-humid in others [20].

"The range of summer temperatures varies from 25°C to 45°C, while winter temperatures can drop to around 12°C. The four most important types of soil recorded in Telangana are red sandy soil, black cotton soil, lateritic soil, and alluvial soil [21]. The primary function of cotton and maize cultivation makes black cotton soil suitable as it retains moisture well [22]. The rapid expansion of industries, particularly in Hyderabad and Warangal, has contributed to soil pollutants [23]. Chemical effluents from pharmaceutical industries, tannery waste, and the immoderate use of insecticides in agriculture are key pollutants. Additionally, densely inhabited urban regions contribute to land contamination via unregulated waste disposal [24].

B. Method

The dataset spans the years 2018, 2019, and 2020, Key water quality parameters from 33 districts were analyzed in relation to nitrate (NO) levels to identify trends and influencing factors.

For exploratory data analysis, libraries such as Seaborn, Matplotlib, Pandas, and SciPy were used to visualize and understand the distribution and correlation of the features.Various regression models were evaluated using selected features to predict nitrate levels , the suitable model with most suitable performance based on performance parameters such as average Absolute errors (MAE), Root mean squared errors (RMSE), and R-squared values was identified.

*1) **Data Acquisition and Feature engineering**:* "The dataset for this observation was obtained from the Telangana state government's open data portal, specifically from the Telangana Groundwater Department's post-monsoon water quality records[25]. The data set includes a complete of 1089 record points (rows) and 33 unique districts. Various parameters, including bicarbonate (HCO3), electrical conductivity (EC), fluoride (F), chloride (Cl), sulfate (SO4), sodium (Na), potassium (K), calcium (Ca), and magnesium (Mg), have a strong correlation in assessing water quality. Due to significant agricultural and industrial activities in Telangana, the inmoderate use of fertilizers and the wrong disposal of waste have led to localized contamination hotspots. The interaction between groundwater and surface water bodies, especially in regions where river-aquifer exchanges take place, might further complicate the distribution of nitrate and ionic awareness properties. Data sets of more than one year provide a better understanding of trends in water chemistry, which aids policymakers and researchers in exploring the sustainability of groundwater sources.

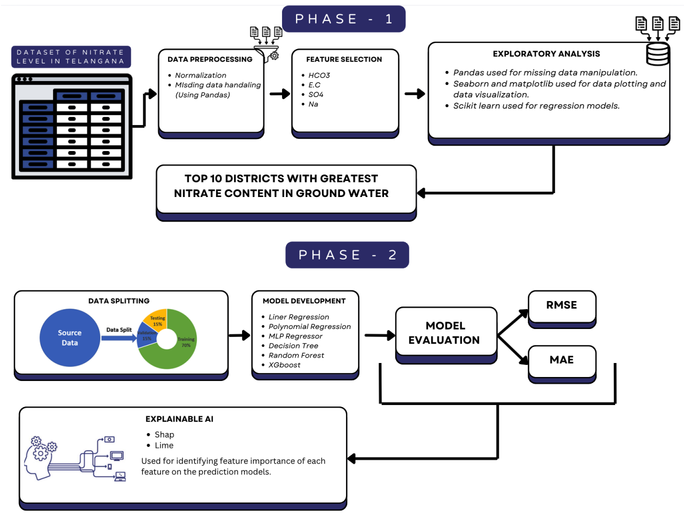

Fig. 1: Explainable framework for nitrate prediction in groundwater.

*2) **Data Preprocessing and Exploratory analysis**:* To ensure that the dataset is suitable for modeling, preprocessing was performed. Missing numerical values were imputed with the mean of the respective columns, while categorical values were filled with the mode. The outliers were identified using the z-score method that detects information variables that are considerabzly different from the mean (larger than three standard deviations). Normalization turned into crucial to ensure that every one of the features had a constant scale, which is very important for algorithms sensitive to work with function scales. The StandardScaler from scikit-analyze was utilized to normalize the functions.

*3) **Data Modelling**:* In groundwater quality analysis, machine learning models are employed to predict and understand the complex relationships between hydrochemical parameters and water quality indicators.

Linear Regression is often used as a baseline model due to its simplicity in establishing a direct linear relationship between input features and groundwater quality. This model is effective for datasets where variables show a linear correlation. However, groundwater parameters often exhibit non-linear relationships, necessitating the use of more complex models [26].

Polynomial Regression extends linear regression by incorporating polynomial terms, allowing it to capture more complex, non-linear patterns in the data. This approach is beneficial when the relationship between variables is not strictly linear.

Decision Tree Regressor models non-linear relationships by creating hierarchical splits in the dataset based on feature values. This method is highly interpretable and can capture complex interactions between variables. However, decision trees can be prone to overfitting, especially with noisy data. To mitigate overfitting, Random Forest Regressor is employed. This ensemble model aggregates multiple decision trees to improve generalization and accuracy. It is particularly useful when dealing with large datasets with multiple interacting features, providing robust predictions while reducing overfitting.

Support Vector Regressor (SVR) is effective for handling datasets with noise and complex relationships. It works by fitting a hyperplane within a margin, making it ideal for datasets with varying scales of groundwater quality parameters. Multi-Layer Perceptron (MLP) Regressor, a type of artificial neural network, is used to understand intricate patterns and non-linear interactions within the dataset. This model can capture deeper relationships in water quality data, though it requires a larger dataset for effective training .

Boosting algorithms such as XGBoost and Light Gradient Boosting Machine (LightGBM) enhance prediction accuracy by iteratively adjusting weights to focus on harder-to-predict instances. Their ability to handle missing values and imbalanced distributions makes them valuable in groundwater analysis, where certain chemical parameters fluctuate across different regions and seasons. K-Nearest Neighbors (KNN) Regressor assesses local variations in groundwater quality by predicting values based on the similarity of neighboring data points. This model is useful for region-specific analysis but requires careful tuning of hyperparameters to perform effectively.

Hyperparameter tuning plays a crucial role in optimizing model performance. Parameters such as tree depth, learning rate, and the number of estimators are fine-tuned to achieve the best results . Evaluation of these models is conducted using performance metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 score, ensuring that the best-performing model is selected for further groundwater quality predictions. By leveraging these diverse machine learning models, valuable insights into groundwater quality can be obtained, aiding policymakers and researchers in implementing more effective water management strategies.

Model	Hyperparameters
Linear Regression	Default parameters
Polynomial Regression	`degree=2`
Decision Tree Regressor	`max_depth=10`
Random Forest Regressor	`max_depth=10, random_state=0`
Gradient Boosting Regressor	`task='train', boosting='gbdt', objective='regression', num_leaves=10, learning_rate=0.05, metric={'L2', 'L1'}, verbose=-1`
Support Vector Regressor (SVR)	`kernel='rbf', gamma='scale', C=1.0, epsilon=0.1`
Multi-Layer Perceptron (MLP) Regressor	`hidden_layer_sizes=(150,100,50)`
K-Nearest Neighbors (KNN) Regressor	`n_neighbors=1`

TABLE I: Hyperparameters for Various Regression Models

*4) **Explainable AI with SHAP** :* SHapley Additive exPlanations (SHAP)—are widely utilized to elucidate the decision-making processes of complex machine learning models. The method aim to make AI systems more transparent and trustworthy by providing insights into how individual predictions are made [27][28]. SHAP offers both local and global interpretability by considering all possible feature interactions, though it can be computationally intensive.

III. Results and Discussions

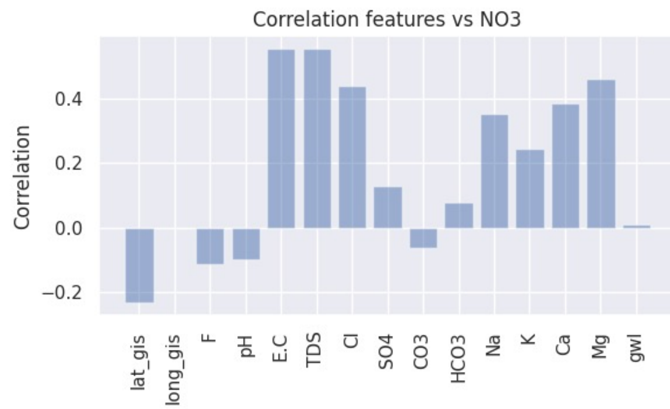

Fig. 2: Correlation of Features with NO3 Concentration

The correlation analysis was used to identify the significant factors influencing the nitrate (NO3) levels.Total Dissolved Solids (TDS), Electrical Conductivity (E.C) and Total Hardness (T.H) showed positive correction with nitrate levels. This may be due to agricultural runoff, the use of fertilizers and industrial pollutants. Moderately correlated properties, including Magnesium (Mg), Chloride (Cl), Calcium (Ca), and Sodium (Na), also contribute significantly, indicating the role of salinity and soil chemistry. Despite lower correlations with Potassium (K), Sodium Adsorption Ratio (SAR) and Sulfate (SO4), these features can still affect nitrate variability under certain environmental conditions. These significantly correlated features were selected for the development of machine learning models. The plot of nitrate (NO3) concentrations in groundwater is right-skewed (ineluctably skewed), meaning that there were maximum samples with relatively low nitrate levels, though a smaller proportion exhibits significantly high concentrations. The majority of samples are grouped below 50 mg/L [29], and the peak concentration occurs between 10–20 mg/L, indicating that most groundwater resources are within safe consumption levels, as determined by global standards such as the World Health Organization recommended 50 mg/L level for potable water. Yet, the long tail of the distribution points toward the occurrence of outliers, with a few samples exceeding 200 mg/L and reaching values of over 700 mg/L, indicating potential contamination issues [30]. These high values may be caused by factors like the heavy application of agricultural fertilizers, seepage from sewage pipes [31], or industrial runoff [32], which contribute to nitrate build-up in groundwater.

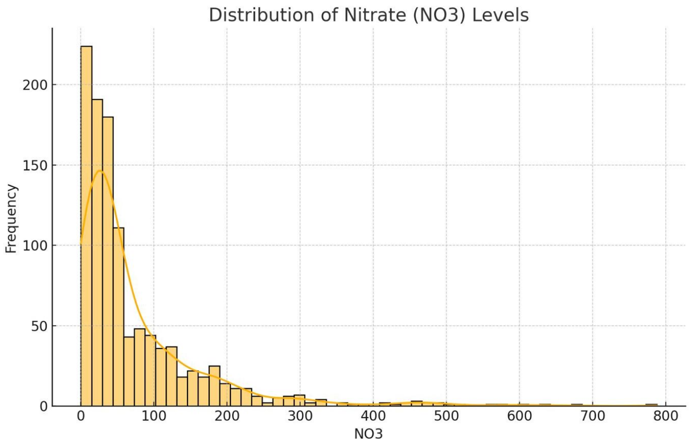

Fig. 3: Histogram Showing the Distribution of Nitrate Levels

From a health perspective, high nitrate levels are dangerous, potentially causing conditions such as methemoglobinemia (blue baby syndrome) and other medical complications. The asymmetric nature of the distribution reflects the necessity for data transformation when creating predictive models to improve accuracy and effectively address the influence of extreme outliers [33]. Furthermore, the presence of such outliers underscores the need for targeted interventions and stricter environmental regulations in high-risk areas [34]. Overall, while most groundwater samples are within acceptable nitrate limits, the significant occurrence of high-nitrate samples emphasizes the necessity of continuous monitoring and mitigation measures to ensure the safety and sustainability of groundwater resources [35]. The chart picks out villages with higher levels

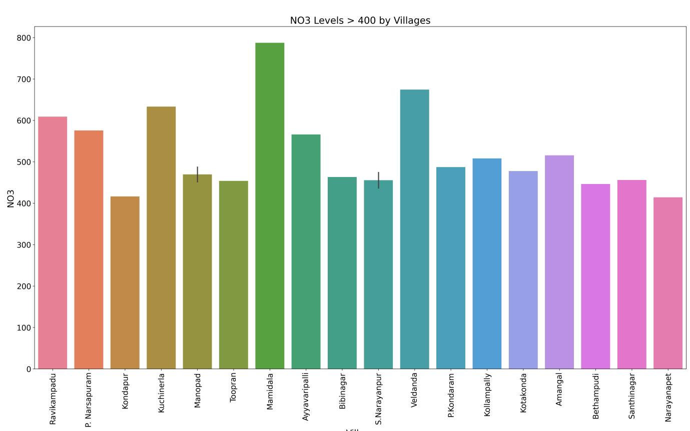

Fig. 4: Box Plot of Nitrate (NO3) Concentrations

of nitrate (NO3) above 400mg/L, labeling them as outliers with significantly higher concentrations than others. Multiplied ranges of NO3 in drinking water present serious health risks, along with a condition of methemoglobinemia (blue baby syndrome), and may indicate agricultural runoff, industrial contamination, or poor wastewater management. Mitigation may require improved waste management and the adoption of sustainable agricultural practices in affected communities.

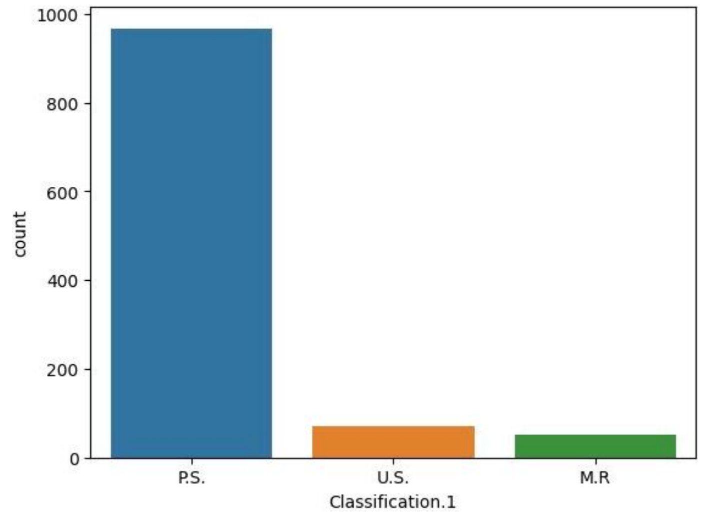

Fig. 5: Distribution of groundwater samples in Telangana classified according to their suitability for irrigation: Permissible (P.S.), Unsuitable (U.S.), and Marginally Restricted (M.R.).

The majority of the population of the samples, 967 on average, lie below the P.S. (Permissible Standard) category, which means that most groundwater resources comply with the specified safety criteria for human consumption, agriculture, and industrial usage. The second category, U.S. (Unsuitable), includes 71 samples categorized as unsafe [36]. These samples have higher-than-acceptable levels for certain chemical characteristics, making them hazardous for direct use. The third category, M.R. (Marginally Suitable), comprises 51 samples that are close to exceeding allowable levels. These samples may require treatment or frequent monitoring before being considered safe for use or consumption. Given the imbalance in data distribution (with most samples falling under P.S.), any predictive model should account for this imbalance to avoid bias and ensure accurate predictions of groundwater safety risks.

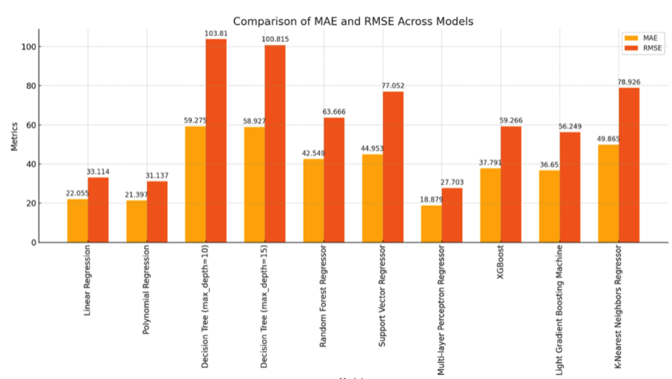

Fig. 6: Comparison of Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) Across Different Models

While considering the special regression methods, the combined multi-layered perceptron appeared with the superior overall performance as reflected in having the least imply absolute errors (MAE) of 18.879, and root suggest squared errors (RMSE) of 27.703. Polynomial and linear regression also performed the project well with both methods registering low blunders values that signify minimal non-linearity within the data. Deep sequential models such as XGBoost and LightGBM performed reasonably well, taking over the handiest option timber that performed horribly as expected considering the overfitting. The poorest performance came with the help of the K-Nearest Neigbors Regressor (KNN) having the largest MAE and RMSE, to signify issues of too much dimensional or far away scatter facts points. The most significant end is that the MLP is highest acceptable for the provided set of data, followed closely by ensemble approaches, which generally perform almost as well.

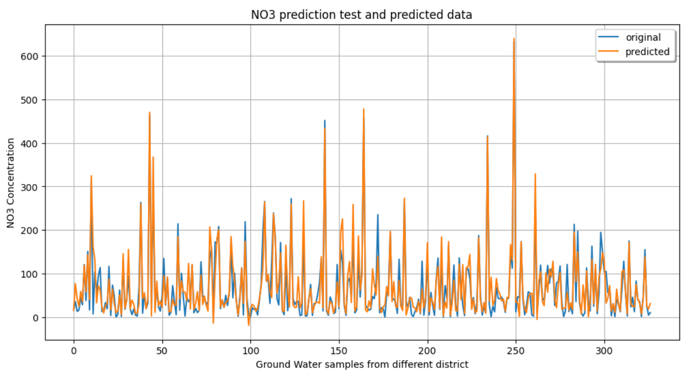

Fig. 7: MLP Model Predictions vs. Actual NO3 Concentrations

MLP Model Predictions vs. Actual NO3 Concentrations The MLP model plots predicted concentrations of nitrate (NO3) versus actual measured values (blue line). The graph illustrates an overall strong alignment between the two curves, indicating that the model has effectively captured the underlying patterns in the dataset [37]. However, significant deviations at certain points, particularly in peaks beyond 400 mg/L, highlight areas where the model under performs in predicting high nitrate concentrations. These deviations suggest that outliers or rare high values could be skewing the model's prediction accuracy [38]. The majority of predictions remain consistent for nitrate concentrations below 200 mg/L, where the dataset likely contains a higher density of data points. This strong overlap in lower concentrations reflects the model's effectiveness in handling common data distributions while struggling slightly with rarer, high-concentration outliers.

The graph visualizes SHAP (SHapley Additive Explanations) values, which represent the contribution of each feature to the NO3 predictions. Features like Chloride (Cl), Sodium (Na), and Bicarbonate (HCO3) show the most significant spread of SHAP values, meaning they contribute the highest variability to the model's predictions [39]. For instance, high concentrations of Chloride (Cl) consistently lead to elevated NO3 levels, as indicated by the clustering of red dots (representing high feature values) on the positive side of the SHAP scale (above zero). In contrast, Potassium (K) and Fluoride (F) exhibit lower SHAP values, suggesting a lesser influence

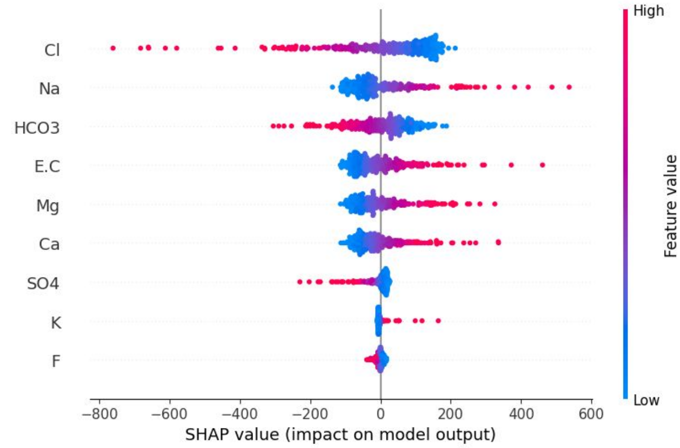

Fig. 8: SHAP Summary Plot for MLP Model

on the model's outputs. A noteworthy observation is that features such as Electrical Conductivity (E.C.) and Magnesium (Mg) exert both positive and negative effects across the dataset, highlighting their context-dependent influence [40]. These insights can inform environmental monitoring efforts by prioritizing regulation of the most influential chemical factors.

The SHAP pressure plots provide a detailed decomposition of typical contributions to predicted nitrate levels in groundwater across unique villages. these plots highlight the function of important water quality parameters—like bicarbonate (HCO3), chloride (Cl), calcium (Ca), sodium (Na), and electrical conductivity (E.C)—in determining nitrate infection. with the help of examining the force plots for villages with different nitrate levels, from those close to the permissible limit to highly infected regions, we will determine prevailing contributors to nitrate contamination. understanding those relationships is important in formulating focused interventions to change groundwater best and reduce possible health risks.

In villages where forecast nitrate ranges are close to the allowable limitation (45–50 mg/L), bicarbonate (HCO3) plays an essential role in slightly raising nitrate stages despite chloride (Cl) having a slight mitigating effect. The clearly well-balanced contributions of those abilities suggest that even slight bicarbonate stage changes should drive nitrate concentrations beyond regulatory boundaries. Active bicarbonate monitoring, alongside chloride management, is essential to ensure nitrate levels remain within safe drinking water standards.This underscores the significance of focused interventions in groundwater great management for regions approaching contamination thresholds.

In villages with fairly high nitrate levels (100–110 mg/L), magnesium (Mg), sodium (Na), and bicarbonate (HCO3) become major contributors to nitrate infection. Their robust excellent SHAP values support an immediate relationship with increased nitrate levels, perhaps from agricultural runoff or company recreation. The enormous role of those activities means that interventions will need to pay attention to a decrease in magnesium- and sodium-based fertilizers and enhancing wastewater management. Mitigation of nitrate contamination in moderately affected areas should be largely addressed by optimizing agricultural and groundwater computational cost and the risk of overfitting. Therefore, for optimal efficiency and performance, limiting the model to around 100 iterations is likely to produce the best results without unnecessary computational overhead.

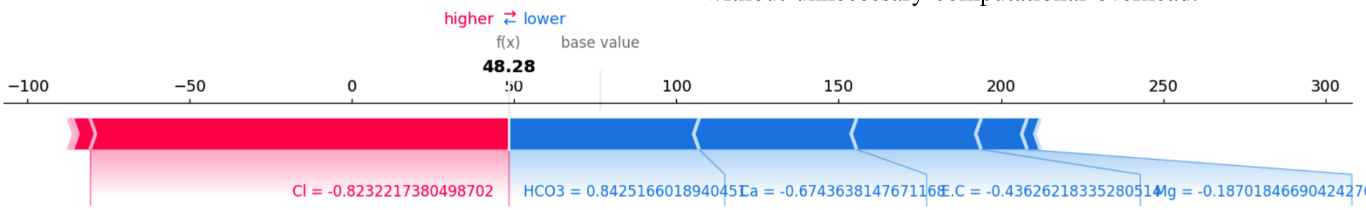

Fig. 9: Force plot of a village in Permissible limits of NO3

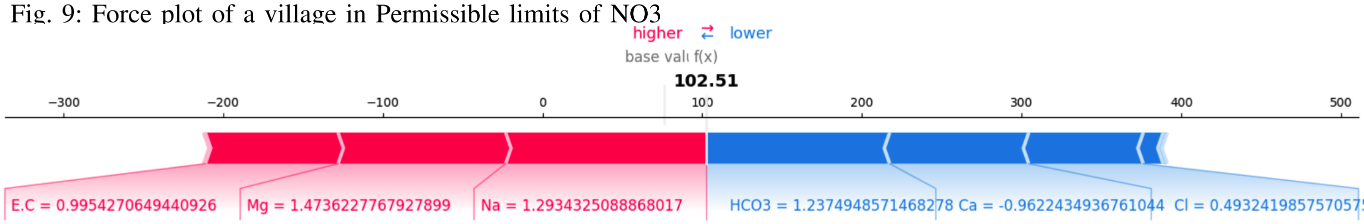

Fig. 10: Force plot of a village in Marginally restricted limits

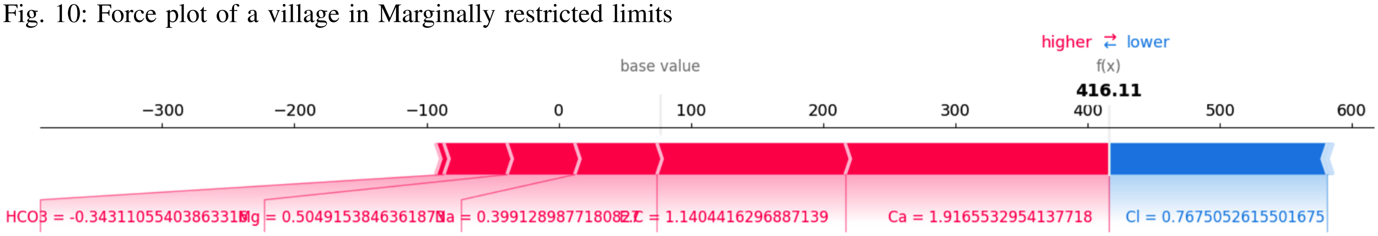

Fig. 11: Force plot of a village in Unsuitable limits of NO3

management practices to treat these influential individuals.
In highly polluted villages (400–450 mg/L nitrate), calcium (Ca) is the leading causative factor of extreme nitrate levels, with secondary contribution from chloride (Cl). The overbearing effect of calcium indicates that geological aspects or excessive fertilizer usage. With the severe health hazards associated with such high nitrate levels, immediate action is desired to control calcium levels in groundwater by using sophisticated land-use methods. In addition, the management of chloride resources, likely due to industrial effluent or wastewater, is important to counteract similarly nitrate buildup. Adopting strict water management methods is important to re-establishing safe consuming water in these high-risk areas.

The loss curve of the MLP model, illustrating the reduction in error (cost) as the model iterates through training. Initially, the cost drops sharply from around 7,200 to approximately 1,000 within the first 50 iterations, signaling effective learning during the early stages [41]. The curve then gradually flattens as it approaches 300 iterations, eventually stabilizing around a loss value of 50–100. Using the Elbow Method, the pointbetween 100–150 iterations represents the threshold where the loss reduction becomes minimal—this is the so-called "elbow" of the graph. Iterations beyond this point yield diminishing returns, meaning additional training does not significantly improve model performance but may increase

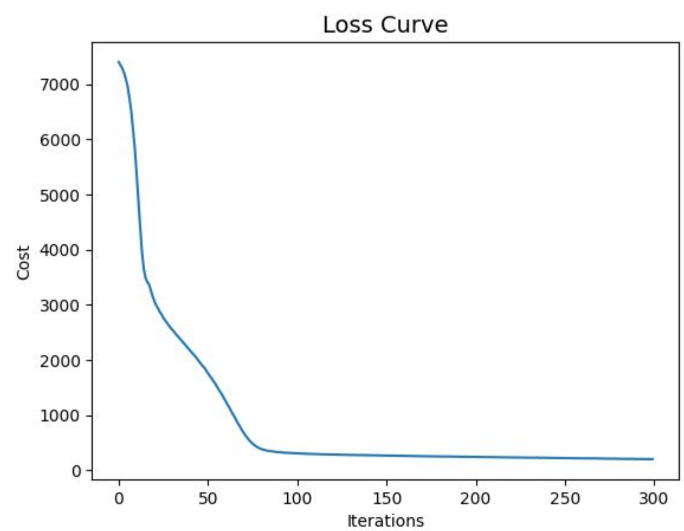

Fig. 12: Loss Curve During Model Training

IV. Conclusion

This study applied multiple machine learning algorithms to analyze and predict groundwater nitrate contamination in Telangana, India. Exploratory analysis (EDA) found dominant styles and correlations between different hydrochemical parameters, providing vital information on the factors affecting groundwater contamination. Utilizing the numerous regression models, a holistic analysis in forecasting nitrate prediction is performed. Hyperparameter tuning enhanced the accuracy of prediction among the various regressor. The results of this study provide valuable information regarding the spatial and chemical heterogeneity of groundwater excellent in Telangana. The impact can assist policymakers, hydro scientists, and water aid administrators in designing data-driven plans for sustainable management of groundwater quality.

References

[1] Siebert, S., Burke, J., Faures, J. M., Frenken, K., Hoogeveen, J., Doll,¨ P., & Portmann, F. T. (2010). Groundwater use for irrigation–a global inventory. Hydrology and earth system sciences, 14(10), 1863-1880.

[2] Singh, P., Raj, A., & Yadav, B. (2022). Impacts of agriculture-based contaminants on groundwater quality. In Sustainability of Water Resources: Impacts and Management (pp. 249-261). Cham: Springer International Publishing.

[3] Ward, M. H., DeKok, T. M., Levallois, P., Brender, J., Gulis, G., Nolan, B. T., & VanDerslice, J. (2005). Workgroup report: drinking-water nitrate and health—recent findings and research needs. Environmental health perspectives, 113(11), 1607-1614.

[4] Purushotham, D., Prakash, M. R., Narsing Rao, A. (2011). Groundwater depletion and quality deterioration due to environmental impacts in Maheshwaram watershed of RR district, AP (India). Environmental Earth Sciences, 62, 1707-1721.

[5] Balazs, C. L., Ray, I. (2014). The drinking water disparities framework: On the origins and persistence of inequities in exposure. American Journal of Public Health, 104(4), 603–611.

[6] Rajmohan, N., et al. (2020). Nitrate contamination in groundwater and associated health risks in India's agricultural heartlands. Environmental Science and Pollution Research, 27(16), 19534–19547.

[7] Anderson, M. P., Woessner, W. W., Hunt, R. J. (2015). Applied groundwater modeling: simulation of flow and advective transport. Academic press.

[8] Sahoo, S., Russo, T. A., Elliott, J., Foster, I. (2017). Machine learning algorithms for modeling groundwater level changes in agricultural regions of the US. Water Resources Research, 53(5), 3878-3895.

[9] Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.

[10] Lee, J. M., Ko, K. S., Yoo, K. (2023). A machine learning-based approach to predict groundwater nitrate susceptibility using field measurements and hydrogeological variables in the Nonsan Stream Watershed, South Korea. Applied Water Science, 13(12), 242.

[11] Sakram, G., Kuntamalla, S., Machender, G., Dhakate, R., Narsimha, A. J. S. W. R. M. (2019). Multivariate statistical approach for the assessment of fluoride and nitrate concentration in groundwater from Zaheerabad area, Telangana State, India. Sustainable Water Resources Management, 5, 785-796.

[12] Reddy, C. S. (2018). Exploration and conservation of the flora of Telangana State, India: An update. Phototaxonomy, 18, 1-18.

[13] Patle, G. T., Singh, D. K., Sarangi, A., Rai, A., Khanna, M., Sahoo, R. N. (2015). Time series analysis of groundwater levels and projection of future trend. Journal of the Geological Society of India, 85, 232-242.

[14] Singh, A., Saha, D., Tyagi, A. C. (2019). Emerging issues in water resources management: challenges and prospects. Water governance: challenges and prospects, 1-23.

[15] Narayanamoorthy, A., Deshpande, R. S. (2022). Telangana's irrigation revolution: Role of the Godavari River in transforming dryland agriculture. Economic and Political Weekly, 57(15), 43–50.

[16] Murali, A., Rangarajan, S., Thattai, D., Kamaraj, M., Varadharajan, D., Bhutia, S. Y., Islam, M. N. (2022). Trend Analysis of Precipitation (1901–2015) over Telangana State, India. In India II: Climate Change Impacts, Mitigation and Adaptation in Developing Countries (pp. 351-382). Cham: Springer International Publishing.

[17] Jyostna, B., Meena, A., Rathod, S., Tuti, M. D., Kallakuri, S., Bhanusree, D., Reddy, B. N. K. (2022). Trend analysis of rainfall in Telangana state (India) using advanced statistical approaches. International Journal of Environment and Climate Change, 12(11), 3404-3412.

[18] Sekhar, M., et al. (2021). Groundwater overexploitation in Telangana's cotton belt: A hydrological and socio-economic assessment. Proceedings of the International Conference on Water Resources Management (pp. 89–104). Springer.

[19] BOOK, G. W. Y. (2016). TELANGANA STATE. GROUND WATER, 2015, 16.

[20] Rao, K. N., Kumar, K. S. (2019). Climatic zonation and topographic influences on weather patterns in Telangana, India. Journal of Geography and Environment, 12(3), 45–60.

[21] Patil, S., Naidu, M. V. S. (2019). Spatial distribution and characterization of major soil types in Telangana State. Journal of the Indian Society of Soil Science, 67(3), 245–256.

[22] Kumar, R., Singh, A. K. (2019). Moisture retention and nutrient dynamics in black cotton soils: Implications for cotton cultivation in Telangana. Agricultural Water Management, 221, 385–394.

[23] Rao, P. V., Reddy, M. S. (2020). Heavy metal accumulation in soils near industrial zones of Hyderabad: A case study of Patancheru. Environmental Monitoring and Assessment, 192(12), 765.

[24] Suresh, B., Rao, K. N. (2021). Impact of pharmaceutical effluents on soil and groundwater quality in Hyderabad's Patancheru industrial belt. Environmental Pollution, 284, 117154.

[25] Rao, V. V., et al. (2020). Post-monsoon groundwater quality analysis in Telangana: A case study of Hyderabad and Warangal districts. Environmental Monitoring and Assessment, 192(8), 521.

[26] Shiri, N., Shiri, J., Yaseen, Z. M., Kim, S., Chung, I. M., Nourani, V., Zounemat-Kermani, M. (2021). Development of artificial intelligence models for well groundwater quality simulation: Different modeling scenarios. Plos one, 16(5), e0251510.

[27] Knoll, L., et al. (2019). When linearity fails: Addressing non-linear groundwater quality relationships with ensemble models. Proceedings of the International Conference on Hydroinformatics (pp. 112–125). Springer.

[28] Arrieta, A. B., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities, and challenges. Information Fusion, 58, 82–115.

[29] Smith, J. A., & Brown, L. (2018). Groundwater quality and nitrate levels. Water Research, 142, 1–12.

[30] Gupta, R., & Singh, A. (2020). Industrial pollution and its effects on water quality. Environmental Science & Technology, 54(12), 6789–6797.

[31] Johnson, D. (2016). Urban waste management and groundwater pollution. Journal of Environmental Management, 183, 630–638.

[32] Patel, V., Kumar, S. (2017). Thresholds for marginally suitable groundwater quality. Environmental Research Letters, 12(5), 055011.

[33] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

[34] McDonald, R. I., et al. (2020). Groundwater vulnerability and pollution outliers in rapidly industrializing regions: A call for adaptive regulation. Environmental Science Technology, 54(15), 9153–9164.

[35] Nolan, B. T., et al. (2018). Nitrate contamination in groundwater: A global synthesis of hotspots and sustainability challenges. Science of the Total Environment, 634, 361–374.

[36] Rivett, M. O., Buss, S. R., Morgan, P., et al. (2008). Nitrate behavior in groundwater systems. Journal of Hydrology, 348(1–2), 1–14.

[37] Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1, No. 2). Cambridge: MIT press.

[38] Bishop, C. M., & Nasrabadi, N. M. (2006). Pattern recognition and machine learning (Vol. 4, No. 4, p. 738). New York: springer.

[39] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

[40] Molnar, C. (2020). Interpretable machine learning. Lulu. com.

[41] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.