International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 21s, 2025
https://theaspd.com/index.php

# Ai-Enhanced Fuzzy-Drastic Model For Groundwater Vulnerability Assessment And Dynamic Contamination Risk Prediction

**Rusul A Al-ameri[1*], Ghufran Ahmed Jawad Al-Baaj[2], Israa Rahman Ghanim[3]**
[1]Najaf Technical Institute, Al-Furat Al-Awsat Technical University, Najaf, Iraq- rusulalameri93@gmail.com
[2]Department of Aviation Techniques, Najaf Technical Institute, Al-Furat Al-Awsat Technical University, Najaf, Iraq- ghufrancivil@atu.edu.iq
[3]Shiite Endowment Diwan, Shiite Endowment Directorate in Najaf, Najaf, Iraq- esraarahman6@gmail.com

***ABSTRACT***
*The groundwater contamination especially in the semi-arid region is a significant threat to public health, ecological balance and sustainable agriculture. Current groundwater vulnerability assessments are usually not temporally sensitive and do not represent complex spatial interactions in different regions and years. The goal of this work is to build an advanced, explainable and data driven framework to predict amount of groundwater contamination risk using geohydrological parameters, chemical water quality characteristics and temporal variations. Unlike conventional models, a hybrid framework composed of Fuzzy-DRASTIC model for uncertainty aware vulnerability indexing and Spatio-Temporal Graph Attention Network (ST-GAT) for intelligent risk classification is introduced in this research. Domain based fuzzification and spatial and temporal attention mechanisms are used in the approach to capture the real-world aquifer dynamic more accurately. Preprocessing of the groundwater quality datasets (2018–2020) by means of normalization and spatial-temporal merging are presented, and the methodology starts from there. Fuzzified geohydrological parameters are used in constructing a Fuzzy DRASTIC Vulnerability Index (FVI). Finally, the final graph structured input that combines chemical attributes, FVI and spatial temporal edges is integrated. This leads to an application of a ST-GAT architecture to model interactions between space and time, and to predict contamination risk categories (Safe, Marginal, Unsafe). Compared to traditional models the proposed Fuzzy-DRASTIC + ST-GAT model has overall accuracy of 98.5%. The model also achieved an excellent separation of the classes as shown by high AUC-ROC score of 0.961. A spatial risk map of high resolution was generated for targeting in high-risk areas. This research provides a high performance, interpretable, and scalable risk prediction solution for groundwater contamination, together in an integrated framework. It allows for proactive water resource management through spatio-temporal deep learning and fuzzy logic combination that enables policy level decision making and sustainable development planning.*

***Keywords:*** *Groundwater Contamination, Fuzzy-DRASTIC Model, Spatio-Temporal Graph Attention Network (ST-GAT), Vulnerability Index, Risk Prediction*

## 1. INTRODUCTION

Groundwater is a vital natural resource that facilitates ecological balance, economic trade, as well as human survival in most parts of the world [1]. In particular, it is a reliable, sustainable source of fresh water in the areas where rivers and lakes are either unavailable or only seasonally available [2]. Irrigation, potable water to rural and urban populations, ground water support industrial operations [3]. It is a mainstay of water security and food sovereignty in many developing countries including India where it contributes more than 60% of irrigated agriculture to up to 85% of rural drinking water supply [4] [5]. Over time, the human activities have gone unchecked resulting in over extraction and poor quality of groundwater leading to dependency. Groundwater is increasingly contaminated because of widespread discharges of untreated or poorly treated waste into the subsurface being caused by rapid urbanization, industrialization, and intensified agricultural practices [6]. They are hazardous substances, such as heavy metals, nitrates, arsenic, fluoride, and pathogens which have a serious health risk and degrade the usability of aquifers [7]. Additionally, contamination of groundwater systems tends to occur underground, a subsurface condition, making early detection costly and remediation so expensive [8]. Environmental

planners and policymakers face a pressing concern over effective groundwater quality management because these challenges are compounded by climate variability, landuse changes, and increased pressure on groundwater recharge zones [9]. The increasing know-how of groundwater systems and the limitations of traditional techniques have made modeling dynamic, intelligent, and spatially adaptive an urgent task [10]. Contemporary environmental problems require that tools are adaptable to uncertainties, variabilities, and large amounts of multi-source data instead of just being based on static assessments requiring [11]. Integration of fuzzy logic provides a more appropriate representation of ambiguous or imprecise environmental variables, which helps in estimating groundwater vulnerability in a more realistic way [12]. Further advances in artificial intelligence (AI) and machine learning (ML) provide powerful means for data-driven risk prediction, thereby allowing the system to learn from historical data patterns and project future contamination threats more accurately [13].

In particular, this work focuses on utilizing sophisticated machine learning methods for predicting groundwater contamination risks when the field lacks historical data or exhibits large spatial and temporal variability. The study introduces a robust framework that integrates the traditional vulnerability assessments with the modern data driven models and combines them into composite model which is more accurate and reliable from the risk prediction of groundwater contamination. The monitoring of the quality of groundwater, both in real time and with precision, is a critical element of environmental protection, resource management and public health, and this study adds primarily to this need. This research makes its contribution by its multi-spatial disciplinary approach that combines geospatial analysis, fuzzy logic and deep learning to address complex environmental challenges. Methodological features permit the combination of chemical parameters, land use data, soil types and hydrogeological features, which make it highly relevant for places where the land use and groundwater recharge patterns are changing. To address this gap, it is further added that the temporal aspect of the model also allows for contamination risk predictions over time, an integral feature for policymakers to predict and plan interventions to protect vulnerable water sources from contamination. The main key contribution of the study was outlined below:

- This study integrates traditional groundwater vulnerability models (such as DRASTIC) with advanced machine learning techniques (ST-GAT) in order to improve their prediction accuracy for contamination risks on both spatial and temporal scales.
- The study introduces an even more nuanced understanding of possible exposures to groundwater by applying fuzzy logic to traditional DRASTIC parameters so as to develop a Fuzzy-DRASTIC Vulnerability Index (FVI) for a better consideration of uncertainty and changes that occur in environmental conditions.
- This leads the model to make Spatio-Temporal Graph Attention Networks capture dynamic changes in groundwater quality over time due to seasonal variations, land-use changes, and hydrological conditions.
- Thus, research results endorse promotion of sustainable groundwater management practices while also yielding in-depth categories of risk with some action-oriented narrow guidelines defining both the short-term inferences and long-term policy planning.

## 2. LITERATURE REVIEW

Groundwater remains vulnerable to contamination due to anthropogenic interference, posing a permanent threat to the sustainability of groundwater resources. Physically based (PB) models are used primarily for groundwater risk assessments but their application becomes computationally impossible when large space and high resolution are required. Machine learning (ML) models have become an internationally accepted alternative to PB models in light of big data. For widespread applicability, an adequate quantity of observations is required for training the ML model, which is very rarely available, especially in rare events like episodic groundwater contamination. This study, [14] addresses the drawbacks of PB and ML models in estimating groundwater well vulnerability to contamination resulting from unconventional oil and gas development (UD) by way of metamodeling, an application hybridizing both model types. The method is demonstrated in northeastern Pennsylvania, where intensive natural gas production from the Marcellus Shale coincides with local communities' dependence on shallow aquifers. Training of the metamodels was done through classifying vulnerability with respect to those

easily computable predictors in a GIS environment. The metamodels were found to be very accurate, with an average out-of-bag error in classification of less than 5%. The most important predictors for accurate metamodel predictions integrated several considerations such as topography, hydrology, and distance from contaminant sources with features such as inverse distance to the nearest upgradient UD source. Maps of predicted vulnerability with violation reports and historical groundwater quality provided further insights into the prevalence of UD contamination in 94 household wells sampled in 2018. While less than 10% of the wells displayed chemical signatures consistent with UD-produced wastewaters, more than 60% were predicted to be located in vulnerable areas. This indicates an increasing likelihood of future contamination events if sufficient protection against contaminant releases is not ensured. These findings underpin the assertion that hybrid physics-informed machine learning models form a solid and highly scalable basis for assessing groundwater contamination threats. Limitations of this study included reliance on rare contamination incidents and a small sample size of wells that might not fully encompass the spatial and temporal variability of groundwater contamination events.

Groundwater pollution, as an issue with remarkable global significance, poses threats to water supplies and ecosystem health. Groundwater vulnerability assessment is imperative for the protection of the human populations and environment. This study, [15] elucidates the adaptation of the traditional DRASTIC method for mapping groundwater vulnerability using a machine learning approach. Such an adaptation entails integrating several tree-based machine learning algorithms with the framework of the model to optimize parameter weights of the DRASTIC model. The adaptation addresses the two most critical limitations which already exist in the literature. First, it makes available to practitioners an evidence-driven counter to the static, aprioristic nature of the original DRASTIC method that fixed ratings and coefficients provided. Second, with machine learning, the approach uses spatial distribution of groundwater contaminants as an avenue to improve the accuracy of the spatial outcomes. Though the machine learning-based approach does not give super great performances according to standard machine learning metrics, it outperformed the traditional DRASTIC model by mapping vulnerability against the field data of actual nitrate concentrations. It is noted that the supervised classification method produced a vulnerability map for which almost 45% of the highly concentrated areas in nitrate content were predicted as highly vulnerable. In comparison, only about 6% of such areas were indicated as highly vulnerable in the original DRASTIC map. The main difference between the two methodologies is that sufficient nitrate data was available to train the machine learning models. It will be concluded from this study that artificial intelligence will give more reliable results with enough data for training, but it cannot dispense with its share of troubles in regards to data quality as well as in the structural nature of the machine learning model itself.

In response to the commissioning of Rooppur Nuclear Power Plant (RNPP) in Ishwardi, Bangladesh, in 2024, increased attention has been raised regarding environmental monitoring, especially concerning water resources management in the areas nearby. However, it is noteworthy that a considerable gap has existed in the literature and environmental datasets pertaining to the initial years of construction for the plant, as not much research was carried out during the initial phase. In this background, [16] the present study was carried out to assess contamination of ground water with potential toxic elements (PTEs) along with health risks for the residents residing close to RNPP from 2014 to 2015. Groundwater samples were collected seasonally during both the dry and wet seasons from nine sampling points. The samples were analyzed for various water quality indicators, including temperature, pH, electrical conductivity, total dissolved solids, total hardness, and PTE concentrations: iron (Fe), manganese (Mn), copper (Cu), lead (Pb), chromium (Cr), cadmium (Cd), and arsenic (As). For assessment, the study adopted the all-new Root Mean Square Water Quality Index (RMS-WQI) with respect to evaluating groundwater PTE contamination levels and then relied on the human health risk assessment model to appreciate the level of human toxicity risks ensuing from exposure to the said elements. The results indicated that there existed a much general tendency for increased concentrations of potential toxic elements (PTEs) during the wet season when compared to the dry season. The concentrations of Fe, Mn, Cd, and As were beyond the threshold limits provided in the drinking water guidelines. In typical RMS-WQI classification, groundwater rated "Fair" in terms of PTE contamination. The non-carcinogenic risk, assessed by the Hazard Index, displayed that nearly about 44% of total samples for adults and 89% for dry seasons while

67% of children and 100% of children respectively for the wet season exceeds the threshold limit of USEPA (HI > 1). The cumulative HI was found to be greater for children compared to adults throughout the study period. On the grounds of carcinogenic risk (CR), these were Cr > As > Cd. Even if the data being used in this study are from 2014-2015, it offers a reference point to future monitoring while mitigating most potential hazardous impacts induced from RNPP. The biggest limitation of this research is that it relies on past data, which may not even represent current conditions of the environment.

This research [17] delves deeply into the multifaceted dynamics of groundwater vulnerability and examines the usually neglected aspects in predicting groundwater vulnerability vis-à-vis the topography, meteorology, socio-economic conditions, land-use, and geology in Bangladesh, which has created a scenario of water stress. The advanced Random Forest (RF) modeling technique conjures insight into a study of the sampled points concentrated strategically at 200 points along the transect for findings in the identification of the extent of significant vulnerability. A considerable part of the land area, about 21% of the area, found at risk consists of important regions such as Rajshahi, Nawabganj, Naogaon, and Dhaka, while regions like Rangpur, Mymensingh, and Barisal comprise an area of 31% with lesser levels of vulnerability. Topographic attributes, specifically aspect, drainage density, and slope, explain 45% of the entire vulnerability and thus are of utmost importance. Population density and industrial pursuits are responsible for 22% of the vulnerability caused by socio-economic factors. The RF model shows a significant score of accuracy above 90%, proving that groundwater dynamics are indeed complicated. The studies have thrown light on such aspects that create a sustainable ground water management strategy by integrating geological, social, and economic factors. While generating a scientific ground for extremely reliable groundwater vulnerability map generation, it introduces a completely new approach where the often-neglected variables are used for the model-building process through machine learning. These findings should help policymakers and urban planners formulate precise and sustainable groundwater management strategies for a resilient water supply for the increasingly populated Bangladesh. The study has certain limitations, such as the temporal scope of the data because it captures vulnerability for a specific period and lack of real-time data, which limits covering immediate changes, even if it contributes widely to the scientific community. The benefits to science from this study are highly significant.

Nitrate contamination investigation of groundwater was conducted using an integrated scheme for groundwater characterization, risk analysis, and a tiered evaluation method for land and surface runoff contaminations. This approach [18] was particularly on soil chemicals with leachants of contaminants into groundwater in the Upper White River Watershed (UWRW) in Indiana. An integrated vulnerability assessment of aquifers was formulated by combining the distributed watershed model (Soil and Water Assessment Tool-SWAT) with machine learning technique, named as Geospatial-Artificial Neural Network (Geo-ANN). Based on models performance metrics, the outcome indicates that integrated assessment approaches were very effective since performance metrics for models as shown in bracket read as (NSE/R2/PBIAS=0.66/0.70/0.07). These suggest that indeed the assessed integrated aquifer vulnerability assessment technique can estimate aquifer vulnerability as shown in this study. In addition, this is a good efficient guide that forms better management decisions of groundwater resources for policymakers and researchers that are involved in groundwater studies. The research notes however some limitations: possible biasness in model prediction from lack of high-resolution data and unexplained uncertainties in parameter calibrations. Such issues may affect the vulnerability assessment accuracies.

Findings of aquifers are fundamental to the protection and management of groundwater resources for contamination. In the present study, [19] artificial intelligence methods and computational optimization algorithms were applied in a way that supports the groundwater contamination vulnerability assessment. Conventional methods such as DRASTIC and its modified indices (ODM) show certain limitations mainly concerning subjectivity-related questions and impreciseness in evaluating nitrate pollution vulnerability of aquifers. To minimize these drawbacks, enhance the drought susceptibility assessment for contaminants' capability, and give an account of the relevant field information, a two-stage approach was carried out. The first stage is perhaps the more novel one, which optimizes the weight of the DRASTIC parameters using Particle Swarm Optimization and Differential Evolution algorithms, thereby yielding two new Vulnerability Indexes based upon the original DRASTIC model formula, ODVI-PSO, and ODVI-DE. The second strategy implemented a Deep Learning Neural Network with respect to both

indices from Strategy-1 as input data. When validated against nitrate contamination levels, the vulnerability index from Strategy-2 based on DLNN algorithm outperformed all other models. In conclusion, the findings demonstrated that the DLNN model under strategy-2 could take advantage of the additional information from the ODVI-PSO and ODVI-DE indices; thus, improving the modeling of aquifer contamination vulnerability. Strategy-2 was thus concluded to have been the best in determining aquifer vulnerability in the study area, especially in the regions where nitrate concentrations were higher than the permissible limits namely mainly in the southern and central part of the area.

The literature review points out that groundwater contamination from anthropogenic activities has emerged as a dire challenge, and conventional modeling techniques are severely limited in their applications for vulnerability indices. Although PB models rely on hydrological processes as their foundation, they can be computationally intensive and thus less suitable for large-scale applications characterized by heightened spatial resolutions. Conversely, ML techniques can effectively map spatial flexibility of a bigger domain; however, they fall short in terms of data availability, particularly for rare contamination occurrences. Hybrid approaches have shown more significant accuracy and practical applications in vulnerability assessments, such as metamodeling in PB and ML methods, as seen in studies in the Marcellus Shale region, where metamodels achieved high classification accuracy and revealed latent contamination risks, despite insignificant observable violations. The AI-informed modifications of the DRASTIC model, most notably through tree-based algorithms such as Random Forest and XGBoost, provide an advancement beyond traditional aprioristic weighting of parameters. The advantage of data-driven weighting is emphasized in these modeling results, which provide a spatial representation closer to actual contaminant distributions, especially for nitrate. In contrast, the ML model's reliability is contingent on the quantum and quality of available data. The application of advanced AI techniques such as Deep Learning Neural Networks (DLNN) and optimization algorithms such as PSO and DE further fine-tuned vulnerability mapping to achieve superior predictive performance, especially if hybridized with conventional indices. Studies on specific regions like Bangladesh and the Upper White River Watershed in the United States exemplify the necessity of including diverse, localized variables into ML models—topography, land use, socio-economic data, meteorological input—thus achieving high predictive accuracy and making substantive recommendations for resource management, despite challenges such as temporal data gaps, outdated datasets, and calibration uncertainties. In totality, the literature justifies the integration of AI into groundwater vulnerability frameworks, advocating for hybrid, data-adaptive, and spatially justified approaches to managing groundwater sustainably against increasing environmental pressure.

## 3. Research Gap

Several critical gaps in groundwater contamination modeling are addressed in this study. Traditionally DRASTIC, and other risk assessment models often do not include temporal variability and/or spatial heterogeneity to reliably model risk [20]. This research uses Fuzzy-DRASTIC model to create a dynamical, time aware, and spatially conscious framework for analyzing groundwater contamination risks by integrating it with Spatio-temporal Graph Attention Networks (ST-GAT). In addition, it also overcomes the restrictions of static data of conventional models in which they do not consider seasonal variations and hydrological fluctuations. Moreover, fuzzy logic is used to deal with uncertainty of environmental data and has a more flexible and understandable prediction of groundwater vulnerability. ST-GAT is further used to improve spatial attention, realize local variability, and complex relationships between land use, soil type, and contamination sources. Additionally, real time prediction abilities are integrated in this study for the purpose of actionable groundwater management. It thereby closes the gap between 'traditional' and 21st century data driven models by facilitating a more accurate and applicable approach for the prediction of groundwater risk.

## 4. MATERIALS AND METHODS

### 4.1 Study Area Description

The data for this study was collected from the Telangana Open Data Portal which has detailed post monsoon groundwater quality reports of diverse districts of Telangana State, India for the years 2018,

2019 and 2020. The data set is made up of the many water sample test results from the various rural and semi urban villages with the information gathered at the district, mandal and village levels. Spatial identifiers (latitude and longitude) are part of each record so it can be easily spatial modelled with GIS for vulnerability assessment purposes. It features 26 key attributes extracted per sample compromising on many chemical parameters including Calcium ($Ca^{2+}$), Magnesium ($Mg^{2+}$), Carbonates ($CO_3^{2-}$), Bicarbonates ($HCO_3^-$), Total Dissolved Solids (TDS), Residual Sodium Carbonate (RSC), Sodium Adsorption Ratio (SAR), Total Hardness, etc. These result features are essential for the judgement of groundwater usability for irrigation, livestock, and drinking and are vital components of making the Fuzzy-DRASTIC model for groundwater vulnerability. There are also two classification labels provided in the datasets — Classification and Classification1 — for classifying groundwater according to salinity and sodium hazard levels on classes such as C1S1 (low salinity/sodium with suitable suitability for all crops) up to C4S4 (very high salinity/sodium generally unsuitable for irrigation). The RSC index is additionally used to provide the impact of carbonates on soil permeability and at the TDS thresholds, the water safety for livestock and poultry is also evaluated. This comprehensive dataset enables integration of the two types of modeling: qualitative (fuzzy logic) and quantitative (AI based) to assess dynamic groundwater contamination risk over space and time [21].

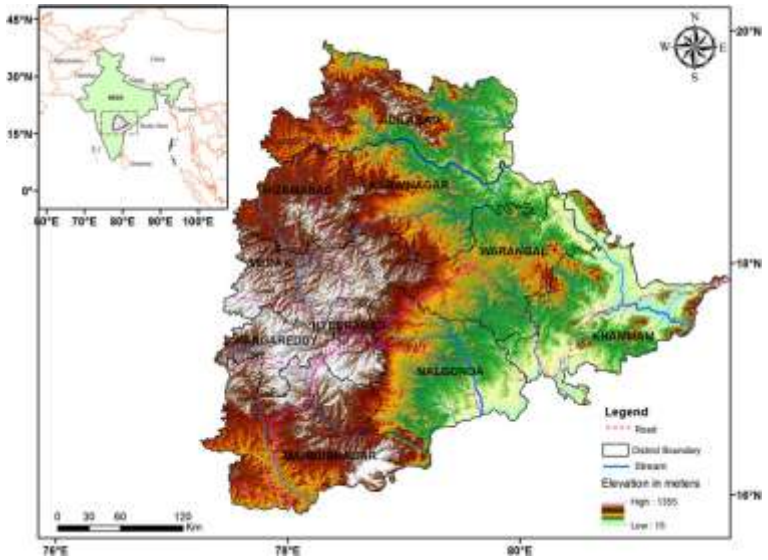


**Figure 1: Location map of the study area (Telangana) India [22].**

### 4.1.1 Data Cleaning

The raw groundwater quality data contained in the three datasets for the years 2018, 2019, and 2020 contain 26 attributes representing various physicochemical parameters of groundwater samples collected across districts of Telangana, India. Mainly includes chemical indicators such as Calcium (Ca), Magnesium (Mg), Carbonates ($CO_3$), Bicarbonates ($HCO_3$), Total hardness, Total Dissolved Solids (TDS), Sodium Adsorption Ratio (SAR), Residual Sodium Carbonate (RSC), and two classification labels. A systematic cleaning operation must be applied to ensure the reliability and quality of input data used for modeling and vulnerability assessment. The first step will deal with any missing values by detecting null entries via the isnull() or NaN checks. Records with missing values in important columns including TDS, SAR, Ca, and Mg, which are important for labeling contamination risks, will be deleted. For non-important fields such as names of villages or classification sub-labels, the missing values will be filled using forward-fill, backward-fill, or median imputation strategies, provided that the missingness is low and not structurally important. Secondly, anomaly detection using the Interquartile Range (IQR) method is performed to identify and remove outliers through visualizing each chemical parameter's distribution. Any TDS value above 10,000 mg/L is suspect and reviewed for potential data entry error or contamination level, along with values of pH which might be recorded to be less than 4 or more than 10, revealing instrument errors or anomalies. Biological or environmental extremes naturally attract domain capping to minimize influence on downstream modeling.

When correcting for datatypes, all chemical concentration fields (such as Ca, Mg, $CO_3$, $HCO_3$, RSC, and SAR) were confirmed to be appropriately stored as floating-point numbers. The geographic coordinates (Latitude and Longitude) would also be cast as floats with appropriate precisions. Furthermore, categorical columns such as District, Mandal, and Classification will be cleaned from leading and trailing white spaces and corrected for inconsistencies in label format. Ultimately, unit standardization will be enforced for keeping consistency among all datasets. The validity of all chemical concentrations is checked, and if found necessary, conversion to a standardized unit is done, such as mg/L or meq/L depending on the parameter and its relevance to the RSC formula or classification of water quality. The concentrations of contributing ions for RSC calculations ($CO_3^{2-}$, $HCO_3^-$, $Ca^{2+}$, $Mg^{2+}$) will be standardized in meq/L using molecular weights and valency to facilitate proper calculation. Such extensive cleaning procedures represent a fundamental element for accurate feature extraction, spatial-temporal merging, and AI groundwater contamination risk evaluation.

**4.1.2 Normalization**

Z-score normalization is the technique applied for all continuous variables in the sampling carried out for groundwater quality to ensure equality of contribution of all input features to the models and the reduced training period for the models to converge. The procedure transforms each feature to a normalized scale by subtracting the mean value and dividing by the standard deviation, as expressed here under in equation:

$$Z = \frac{X - \mu}{\sigma}$$

where $X$ is the individual feature value, $\mu$ is the mean, and $\sigma$ is the standard deviation of the feature. This transformation is to transform different normalized features to mean 0 and standard deviation 1. As these chemical parameters — Total Dissolved Solids (TDS), Sodium Adsorption Ratio (SAR), Residual Sodium Carbonate (RSC), Total Hardness, and ion concentrations such as Calcium (Ca), Magnesium (Mg), Carbonates ($CO_3$), and Bicarbonates ($HCO_3$), follow a wide range of units and ranges, it is particularly suitable for this dataset to use the type of normalization called Z Score. For example, TDS values tend to span much of the range for other parameters, and if not normalized, may tend to unduly affect the training of the model. After applying z_score normalization to these important features, the dataset is standardized for the benefit of AI algorithms, especially linear classifiers such as distance classifiers and gradient based optimizers. This step is of great importance in this step for robust, unbiased contamination risk prediction and accuracy in downstream modeling stages.

**4.1.3 Spatial-Temporal Merging**

All the quality data of groundwater for the future ST-GAT modeling requires data to have the spatial as well as the temporal dimensions structured in a machine-readable format. It starts by merging the three yearly datasets-2018, 2019, and 2020-into a new dataset with an attribute representing the year when each record was made.It resolves column mismatch across the datasets such that all features correctly align before merging. Next, every record is assigned some temporal identifier, which could either be a string that records the district and the year (e.g., "Hyderabad_2019") or a straightforward index for the time line (e.g., 0 for 2018, 1 for 2019, and 2 for 2020). With this temporal tagging, the model will be able to make clear traces in year-to-year variations, thereby establishing the time line for graph learning that the time series need.

The coordinates of the sampling site are given in spatial terms based on the latitude and longitude values. Usually, they are rounded to a predetermined number of decimal places (often 3) to avoid such noise, and group together from samples obtained from such neighbor locations. A unique spatial ID for every location is generated based on its district, mandal, and village information. This ID is useful for generating the spatial adjacency matrix that defines the connection between neighboring sites in the spatial graph. The structure of the final dataset consists of the following components: temporal attributes (year, post-monsoon season), spatial data (latitude, longitude, spatial ID), a normalized set of chemical parameters (TDS, SAR, RSC, Ca, Mg, $CO_3$, $HCO_3$, and total hardness), and labels indicating irrigation suitability (Classification1), safety classes based on RSC, and water usability classes for livestock based on TDS. This

well-structured spatiotemporal dataset forms the basis for further deployment of AI-based models like ST-GAT for the dynamic prediction of groundwater vulnerability and contamination risk in space and time.

**4.2 System Architecture**

The proposed methodology of integrating hydrogeological model and developing powerful AI techniques for the purpose of groundwater contamination risk prediction is proposed. The Fuzzy-DRASTIC model is initially used to evaluate vulnerability of aquifers by means of fuzzy membership functions of the Fuzzy DRASTIC parameters (e.g., depth to water, net recharge, soil media, hydraulic conductivity). The first are then aggregated using fuzzy inference rules to yielding Fuzzy Vulnerability Index (FVI). Also, the groundwater quality indicators such as TDS, SAR and RSC are classified into contamination risk levels. It is structured as a graph where the spatial and temporal components of the data are structured across the nodes, each representing a geographic unit (village or grid cell) and the edges specify spatial proximity and temporal continuity across years. It is provided with a Spatio-Temporal Graph Attention Network (ST-GAT), which employs graph attention (spatial heterogeneity) and temporal attention (time dependent pattern) layers to represent spatial heterogeneity and to learn time dependent contamination patterns. Based on Python implementation, the model shows high accuracy in the outputs, and it provides a robust, interpretable and dynamic framework for the real-world groundwater management and policy making.

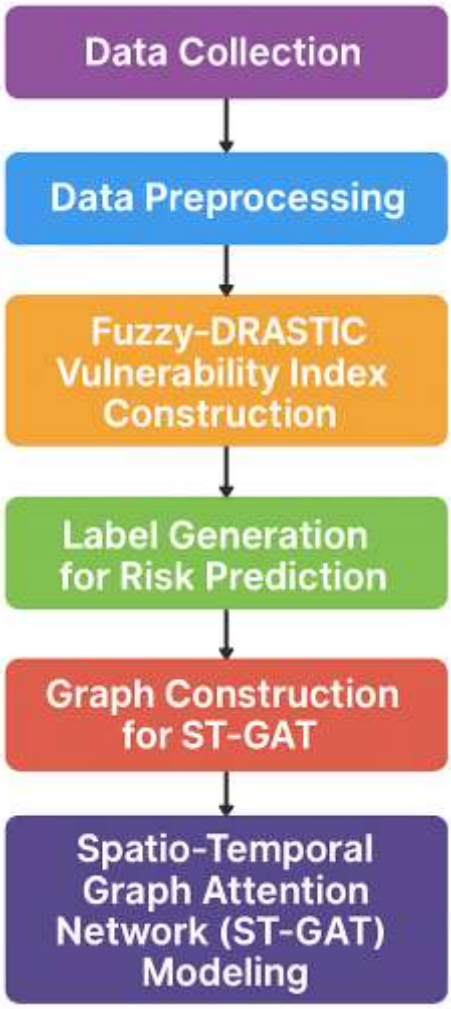


**Figure 2: Workflow of the Proposed Approach**

**4.2.1DRASTIC Model Overview**

The DRASTIC model is an established methodology developed by the United States Environmental Protection Agency to assess groundwater vulnerability against contamination. It takes into consideration

hydrogeological and environmental parameters to obtain a vulnerability index for a particular region. Each letter in DRASTIC stands for one of seven used parameters in the model, each of them being assigned a weight and a rating based on the importance for susceptibility to pollution. This might stand to reason since in the context of Telangana groundwater quality data set (2018-2020) the actual direct measurement DRASTIC parameters may not be entirely complete, but the model favors proxies, additional datasets, and derived values. As below is an explanation for each parameter and the methodology of mapping or how it can be developed from the dataset or through the integration of other external geospatial and hydrological datasets:

- **Depth to Water (D)**

It is a vertical distance from the land surface to the groundwater table, which is a fundamental indicator of groundwater vulnerability. Shallow water tables allow contaminants to easily pollute the aquifer and increase the risk of pollution. The Telangana groundwater dataset (2018-2020) does not provide direct water table depth measurements, but this parameter can be supplemented using regional data from either the Central Ground Water Board (CGWB) or satellite-derived water level models. Using the spatial coordinates supplied (latitude and longitude), points at assessment sites will have interpolated depth values that will yield approximate vulnerability ratings for that specific site. The generally low depth to groundwater will yield a high score in the DRASTIC index, which indicates increased susceptibility to surface-based contamination.

**Net Recharge (R)**

The net recharge represents the volume of water that moves down into the ground and serves to replenish the aquifer system. Indeed, recharge plays a dual role in that while it is an essential factor for aquifer sustainability, excessive recharge may speed up the movement of pollutants due to percolation. Although recharge values do not form a part of the groundwater dataset, these can be computed from annual rainfall data provided by the Indian Meteorological Department (IMD) and soil infiltration rates based on land use. These remote sensing data taking into account such satellite systems as NASA's GPM or TRMM would also help with regional recharge estimations. The higher recharge areas, especially in the post monsoon seasons taken into account in the dataset, may be having higher vulnerability scores because the pollutants travel with a higher potential through infiltration.

**Aquifer Media (A)**

Aquifer media are the geological formations, such as gravel, sandstone, or basalt, which store and transmit groundwater. Permeability and porosity influence the flow rates of contaminants through these materials directly. While this information cannot be obtained from the groundwater quality dataset, aquifer media characteristics may be gleaned from Geological Survey of India (GSI) maps. The common aquifer types in Telangana are fractured granites and basaltic rocks, with each having different levels of permeability. This parameter will have to be integrated into the model by assigning vulnerability scores based on the capacity of different materials to allow pollutant migration, and therefore sand and gravel-typified coarse-textured media would attract higher ratings compared to other media such as loam or clay.

**Soil Media (S)**

Soil mediums are the vertically above weathered zone, which accentuates the significance that allows filtration and percolation of contaminants into the subsurface. Permeability, in greater extents, adsorbs sandy and loamy soils that then encourage pollutants' prompt migration into the subsurface. However, information regarding soil types is not available in the primary files meant for groundwater quality; it can be preferably integrated from soil village maps or national repositories like the National Bureau of Soil Survey and Land Use Planning (NBSS&LUP). The location metadata (district, mandal, village) could assist in matching soil types to the sampling points. Fine-textured soils like clay are poorly permeable (low vulnerability scores) and coarser soils are rated higher as they allow faster infiltration.

**Topography (T)**
Topography denotes the inclination and height of land surface. It helps decide if water accretes and infiltrates into the ground or quickly runs off to carry pollutants. Water enters the ground more on flat terrain, whereas steep slopes favour runoff, thereby diminishing the awareness to recharge groundwater. This parameter does not directly appear in the dataset, but can be derived from Digital Elevation Models (DEMs), i.e. DEMS derived from either SRTM or ASTER. Using GPS coordinates, slope will be calculated for each sample location, and this parameter will then be introduced into the DRASTIC model. Flatter terrains will be characterized by greater vulnerability scores for pollution accumulation and percolation.

**Impact of Vadose Zone (I)**
Vadose zone refers to the area between the surface of the land and the water table. For instance, it acts as filter through which the entire surface water passes through and thereupon as inputs to the aquifer. The inherent properties of such vadose zone material primarily govern contaminant attenuation. This data may be indirectly inferred from the regional geological understanding and borehole logs, same as aquifer media. In Telangana, regions with fractured or weathered rock Vedas under looser vadose zones are associated with fast contamination transport during monsoonal recharge periods. These ratings are assigning for model integration on property permeability-retention characteristics of vadose materials, the more permeable areas are hypothesized to be most vulnerable.

**Hydraulic Conductivity (C)**
Hydraulic conductivity indicates how quickly water can move through aquifer materials. A high conductivity means that once a contaminant penetrates, it can be removed far from source. This parameter is fundamental in understanding susceptibility and spread of contamination. A parameter is not recorded in the dataset; however, it could be derived from CGWB wells or assigned from known conductivity ranges for specific regional rock types. Creating a high DRASTIC score therefore using high conductivity formations like those in sandy aquifers or fractured rocks indirectly represented greater vulnerability to contaminant transport.
The finalization of the DRASTIC model is actually much more exhaustive in the assessment of groundwater vulnerability as compared to what the seven hydrogeological parameters normally stipulate. Indeed, the data available on the groundwater of Telangana, in that respect, provides the required chemical indicators for water quality. However, other spatial, geological, and hydrological inputs would be needed to complete the DRASTIC parameters in this case. With the help of GIS tools and government datasets, the combination of each of the parameters will have to realize in a quantifiable format first which, thus, can produce a spatially explicit vulnerability index that in turn would support proactive management of water resources, policy formulation, and the adoption of precision irrigation and contamination mitigation measures. These parameters can be spatially mapped, quantified, and combined to form a comprehensive vulnerability assessment for irrigation and drinking purposes-the basis for eventual integration with AI models such as ST-GAT for dynamic prediction of groundwater risks.

**4.2.2. Fuzzy-DRASTIC Vulnerability Index Construction**
The DRASTIC model has traditionally provided the framework within the context of these seven critical hydrogeological parameters for groundwater vulnerability assessment, that is, Depth to Water (D), Net Recharge (R), Aquifer Media (A), Soil Media (S), Topography (T), Impact of the Vadose Zone (I) and Hydraulic Conductivity (C). But environmental systems are inherently uncertain and complex, and crisp classifications often fall short. To address this problem the Fuzzy DRASTIC model integrates fuzzy logic idea to more realistically, flexibly, and interpretably assess groundwater vulnerability, by incorporating uncertainties in parameter measurement and the regional variability.

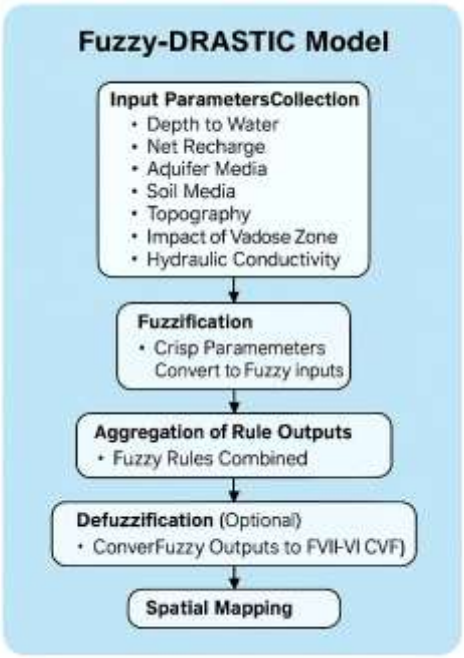


**Figure 3: Fuzzy-DRASTIC Moel**

**Parameter Weighting**

In the Fuzzy-DRASTIC model, the importance of each parameter is first attributed using knowledge in the domain and expert judgment. The classical DRASTIC models assign fixed weights such as D = 5, R = 4, A = 3, etc., considering their relative influence on groundwater contamination. Assigning fuzzy weights is still done according to this as base guidance; however, it can be modified depending on the local hydrogeological setting, especially in a state like Telangana where aquifer characteristics and recharge patterns are spatially heterogeneous. The weights are used to prioritize the parameters that aggregate their effects in the fuzzy inference stage.

**Fuzzification of Parameters**

The process of fuzzification deals with transforming exact quantitative values of DRASTIC parameters into qualitative linguistic terms such as Low, Medium, or High vulnerability. Each of the parameter is associated with a fuzzy membership function, either triangular or trapezoidal or even Gaussian-shaped, and defining the degree to which the respective parameter belongs to a certain category. Membership functions were built done through historical data, knowledge from experts, and environmental benchmarks so as to better capture the gradual transitions between levels of vulnerability. For example:

- **Depth to Water**: Shallow depths (e.g., <5m) will have high membership in the "High Vulnerability" class, while deeper levels (>20m) will fall under "Low Vulnerability".
- **Net Recharge**: Higher recharge rates will have greater membership in "High" due to the increased potential for contaminants to percolate.
- **Soil Media**: Sandy soils may be classified with high vulnerability membership, while clay-rich soils may fall under "Low".

**Rule-Based Aggregation**

The next step after fuzzifying each of the seven parameters-Water Depth, Net Recharge, Aquifer Media, Soil Media, Topography, Impact of the Vadose Zone, and Hydraulic conductivity-into linguistic categories (for instance, Low, Medium & High) is rule-based aggregation. This is a significant operation that reinforces the conversion of qualitative expert knowledge into a quantitative index useful in spatial decision-making.

**1. Rule Formulation**

This fuzzy IF-THEN rule-based system simulates expert reasoning and captures relationships among different combinations of vulnerability factors. These rules are primarily based on hydrogeological knowledge and past studies. A few standard examples are:

- Rule 1: IF Depth is Low AND Recharge is High AND Soil is Sandy, THEN Vulnerability is High.
- Rule 2: IF Depth is Moderate AND Soil is Loamy AND Topography is Flat, THEN Vulnerability is Moderate.
- Rule 3: IF Depth is High AND Recharge is Low AND Soil is Clayey, THEN Vulnerability is Low.

Each rule represents an expert interpretation of how combinations of physical factors influence the potential for groundwater contamination.

**2. Fuzzy Operators (AND, OR, NOT)**

To evaluate these rules, fuzzy logic operators are used:

- Fuzzy AND (minimum operator): Returns the minimum membership value among the input parameters.

Example:

- Depth (Low) = 0.8
- Recharge (High) = 0.6
- Soil (Sandy) = 0.9⟶ Fuzzy AND = min(0.8, 0.6, 0.9) = 0.6

- Fuzzy OR (maximum operator): Returns the maximum value.
- Fuzzy NOT (negation): Converts a membership value μμμ to 1−μ1 - μ1−μ.

These operations help determine how well a particular spatial unit satisfies each rule.

**4. Aggregation of Rules**

Once activated, rules may work together in more than one spatial location, and their outputs now get aggregated. The aggregation process entails the union (maximum membership value) of all activated vulnerability outputs for each class. The Low vulnerability is the maximum of all Low fuzzy outputs: 0.2 and 0.1, so here the maximum value is 0.2. The same goes for the Moderate and High vulnerability maximum aggregations to determine the fuzzy output: e.g., 0.4 and 0.5 for Moderate, taking the maximum gives 0.5; for High, 0.6 and 0.7 maximum is 0.7. Hence, after the aggregation process, a fuzzy vulnerability profile is drawn for each spatial unit, summing up the mutual effect of all considered factors.

**5. Defuzzification**

Next, defuzzification is implemented to transform the fuzzy output set into a particular crisp value, developing particular applications like vulnerability mapping or integration with machine learning. The most widely-used method is centroid defuzzification, the method used to find the gravity centre of the output membership function. This gives the value of FVI or Fuzzy Vulnerability Index on a continuous scale, whereby a value of 0 means the lowest vulnerability and a value of 1, the highest.

**6. Spatial Assignment of FVI**

The computed fuzzy vulnerability index (FVI) values were assigned to the mapping units known as villages, mandals, or grid cells within the Geographic Information System (GIS) for deriving a fuzzy risk map having spatial locations classified into different zones of vulnerability as Low Vulnerability (FVI < 0.3), Moderate Vulnerability (0.3 ≤ FVI < 0.6), and High Vulnerability (FVI ≥ 0.6), such a map serves as a powerful visualization tool enabling planners and decision-makers to pinpoint groundwater zones at various levels of risk along with the confidence level of each zone represented through the FVI score. Rule-based aggregation in the Fuzzy-DRASTIC model is a powerful bridge used to cross over from numerical data to expert knowledge. The hydrogeological intuition is converted into systematic rules capturing complex interactions among multiple fuzzy parameters effects and produces a nuanced vulnerability score. Instead of having rigid classification, it respects the uncertainty and gradients typical

of environmental systems, thus making it appropriate for groundwater quality and pollution risk assessment in highly complex regions like Telangana.

### 4.3 Label Generation for Risk Prediction

**Contamination Classification:**

Contamination classification consists of assigning the groundwater samples under investigation into classes based on chemical characteristics of the water, chiefly determination of status using the existing dataset's 'Classification 1' labels. These labels range from C1S1 to C4S4 and give the different qualities of water based on concentrations of Na and S. For example, C1S1 refers to low salinity and low sodium water, suitable for most kinds of irrigation. C4S4 indicates water with high salinity and high sodium levels, which is generally unsuitable for irrigation unless properly managed. RSC values are paramount in contamination classification, particularly in the assessment of water suitability for irrigation depending on the balance of alkaline ions and earth ions. Using these labels will enable the attribution of a class to each sample representing its contamination status and suitability for irrigation, livestock consumption, or general consumption. Moreover, RSC criteria are critical for the classification because any sample having an RSC value of greater than 2.5 is considered unsuitable for irrigation due to soil impairment such as decreased permeability. RSC values from 1.25 to 2.5 are marginal, while RSC values below 1.25 are suitable for irrigation. This classification provides a base for the identification of contaminated zones from which the dataset strongly points on the identification of which groundwater samples can achieve the quality criteria for different uses.

**Risk Labeling:**

The process of risk labeling goes beyond mere classification in assigning categories based on TDS, RSC, and SAR values. These parameters serve as criteria to assess possible risks to the quality of water in the environment. For example: TDS refers to the total concentration of dissolved substances in water, and on the basis of TDS, water may be classified as Safe, Marginal, or Unsafe for drinking or irrigation:

- Safe: TDS less than 1000 mg/L, typically suitable for both human consumption and irrigation.
- Marginal: TDS between 1000-3000 mg/L, still usable, but may cause some mild health or crop issues.
- Unsafe: TDS above 3000 mg/L, indicating water is potentially harmful for agricultural or human use.

Besides the TDS, the RSC values are also utilized to classify the risk of water for irrigation purposes. Water showing the RSC value more than 2.5 is unsafe for irrigation, while between 1.25 and 2.5 is marginal and must be managed precisely for sustainable use. SAR is also important, as it measures sodium in relation to calcium and magnesium, and high levels of it indicate that water may be unsuited for crops because of the sodium accumulation in the soil. Such supervised learning domains of machine learning models also require TDS, RSC, or SAR classifications as dependent variables. Consequently, this will help train the model to predict groundwater contamination risk for several regions. The risk, which is labeled Safe-Marginal or Unsafe, will be the output variable, aiding modelers to understand how different chemical properties affect pollution risk and how to use this information to develop effective strategies for managing water quality. These labels also sponsor identification of at-risk areas, as well as areas requiring immediate address in groundwater quality improvements.

### 4.4 Spatio-Temporal Graph Attention Network (ST-GAT) Modeling

The graph-structured inputs are the main pillars of the ST-GAT model that reflect the spatial and temporal dimensions of the groundwater monitoring data. Each node in the graph represents some spatial unit, like a village or boa location, or a defined grid cell. The node features comprise normalized chemical properties: Total Dissolved Solids, Sodium Adsorption Ratio, and Residual Sodium Carbonate, alongside hydrogeological scores based on the DRASTIC model and some contextual features, such as land-use patterns, soil classification, etc. The types of edges that the graph will have been: 1. spatial and 2. temporal. The spatial edges are based on proximity, using methods like k-nearest neighbor (KNN) or threshold distance to connect locations with similar environmental- or geographical-scale features. Temporal edges, in contrast, connect the same spatial node across different years, e.g., 2018, 2019, 2020; thus, the model

can learn about change over time. Such a rich structure allows ST-GAT to model very complex relationships across both spatial and temporal domains.

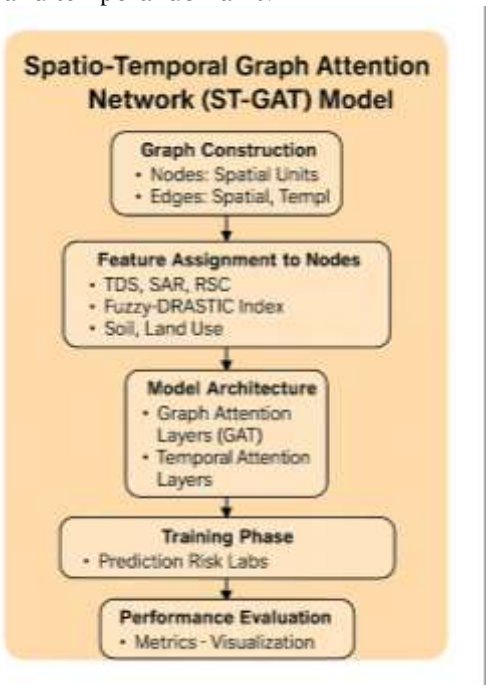


**Figure 4: ST-GAT Model**

**1. Graph Attention Layers (GAT) – Spatial Component**

The temporal component of the ST-GAT architecture utilizes temporal modules, the spatial one finds its representation through Graph Attention Layers (GAT) for portraying the intricacies of spatial interactions across several geographical locations. In a graph structure, a node signifies a spatial unit such as a village, grid cell, or water sampling location. The basic innovation of GAT is that for each neighboring node, it produces dynamically tuned attention weights instead of the same weights for all neighbors. In the real environment modeling, regions adjacent to each other can differ dramatically in terms of their geology, land use patterns, aquifer properties, and sources of contamination. It thus acts in consideration of these significant differences while training the entire model. This model learns which neighbor has more relevance by calculating the attention coefficients that determine the extent to which information coming from each neighbor effectively contributes towards the final representation of the node. For example, imagine two close sites residing adjacent but having an industrial discharge and the other is agricultural; once again, the model efficiently prioritizes the former to foretell contamination in groundwater. Thus, a spatial attention mechanism permits the model to learn fine-scale and indeed heterogeneous geography reflecting local patterns or fine differences in the environment that traditional modeling would hardly capture.

**2. Temporal Attention Layers – Temporal Component**

To add to the spatial learning, the ST-GAT Architecture includes temporal attention layers, which represent a degree of evolution of groundwater quality over time. These layers help the model in learning how historical observations would influence the current state of a particular location. While traditional time series models assume fixed temporal influence on the data, such as say the previous year's value affecting the current year's one with equal importance, an attention mechanism would allow the model to dynamically learn the relative importance of each past timestep. This is specifically useful for groundwater systems in which it takes time for contamination effects to manifest, such as an application of fertilizers in 2018 which could have substantial consequences in terms of groundwater nitrate levels in 2019 or 2020. Temporal attention thus accounts for time-lagged dependencies, seasonal relevant events (e.g., dilution during monsoon or concentration in dry months), as well as cyclic ones. In so doing, it

allows the model to develop a comprehensive temporal perspective on the dynamics of groundwater contamination and enhances prediction accuracy to be context-sensitive.

Combining both Graph Attention Layers and Temporal Attention Layers transforms the ST-GAT into a spatio-temporal modeling framework in which virtually nothing falls short in accounting for inter-regional interactions and time-evolving behaviours such that the real-world complexities of groundwater contamination could be modelled more closely. The spatial aspect concerns itself with where contamination patterns are modulated based on neighboring conditions, whereas the temporal aspect lays emphasis on the when and how that past conditions influence future contamination levels. This dual learning mechanism is especially strong for environmental and hydrogeological systems, whose very existence is spatial and temporal. It is capable of providing highly and finely tuned predictions along with context-sensitive predictions for groundwater vulnerability and contamination risk assessment, which itself would serve as a very important input for environmental planners, researchers, and policy-makers.

**Algorithm 1:**
**Input:**
GW_Quality_Data_2018, GW_Quality_Data_2019, GW_Quality_Data_2020
DRASTIC_Params (Depth, Recharge, Aquifer, Soil, Topography, Vadose, Conductivity)
Spatial_Info (Latitude, Longitude)
Soil_LandUse_Data (optional)
**Output:**
Fuzzy Vulnerability Index (FVI)
Contamination Risk Labels
Risk Prediction Map
**Begin:**
**Step 1. Data Preprocessing:**
For each year in [2018, 2019, 2020]:
Load dataset
Handle missing values:
If critical value missing → drop row
Else → apply forward/backward fill or median imputation
Detect and remove outliers using IQR
Normalize continuous features using Z-Score
Assign 'Year' column
Round coordinates and assign Spatial_ID ← hash(District + Mandal + Village)
Merge all yearly datasets vertically
**Step 2. Fuzzy-DRASTIC Index Computation:**
For each spatial unit:
For each DRASTIC parameter:
Assign rating and weight (based on domain knowledge)
Fuzzify parameter into (Low, Medium, High) membership
Apply fuzzy IF-THEN rules:
e.g., IF Depth is Low AND Recharge is High → Vulnerability is High
Aggregate rules using fuzzy logic operators (max, min)
Defuzzify output (Centroid Method) → Compute FVI
Assign vulnerability class:
If FVI < 0.3 → Low
If 0.3 ≤ FVI < 0.6 → Moderate
If FVI ≥ 0.6 → High
**Step 3. Label Generation:**
For each record:
Use Classification1 + thresholds of TDS, SAR, RSC
Assign Risk_Label ← {Safe, Marginal, Unsafe}

Encode label for model training
Step 4. Graph Construction for ST-GAT:
Define Node ← each unique spatial unit
Connect nodes using spatial proximity (KNN or distance threshold)
Add temporal edges across years for same location
Define Node_Features ← [Normalized chemicals, FVI, land use, soil]
**Step 5. ST-GAT Model Training:**
Initialize ST-GAT Model:
Graph Attention Layer (GAT) for spatial edges
Temporal Attention Layer for time links
Input: Graph structure, Node features
Target: Risk_Label
Train model using cross-entropy loss
**Step 6. Prediction & Visualization**:
For each test node:
Predict risk class ← ST-GAT output
Use GIS tools:
Visualize risk zones (Safe, Marginal, Unsafe)
Generate FVI maps and overlay risk levels
End

## 5. Results and Discussion

In this section, it presents the results obtained from the proposed groundwater contamination prediction risk framework that has been implemented in Python. The spatio-temporal modeling in this thesis includes the integration of groundwater quality data analysis, fuzzy logic based DRASTIC vulnerability assessment and Graph Neural Networks. The results are structured such that key stages of analysis are shown with descriptive statistics for chemical parameters, the generation of the Fuzzy Vulnerability Index (FVI), and the amounts of predictive outputs of the Spatio Temporal Graph Attention Network (ST-GAT) and the quantitative evaluation of model performance using standard classification metrics. The proposed model is evaluated based on its predictive accuracy, regional contamination trends, and vulnerability patterns and these findings are useful in understanding regional contamination patterns.

### 5.1 Experimental Outcome

Figure 5 presents the feature compressions scores from the taught model ST-GAT for the entire project along with the implications of these scores, usually implying some contribution to groundwater contamination risk prediction by each input feature. Total dissolved solids (TDS) outrank all other features with importance score of 0.212. It is a strong input that can determine the contamination level. The Fuzzy-DRASTIC Vulnerability Index vulgarly follows closely with an input score of 0.184, indicating the value of adding hydrogeological and environmental vulnerability in the prediction model. Besides these, SAR and depth to water table also come to the scene as score holders of 0.143 and 0.108, respectively, hinting at their contributions in determining the usability of groundwater and pollutant mobility. Moderately important are parameters like soil media type, net recharge, and residual sodium carbonate, leveling up their roles in groundwater chemical behavior and flow. Somewhat interestingly, in contrast to features such as land use category, these attributes proved to be relevant as well, but most importantly rather weakly influenced (0.014) when compared to those used with hydrochemical indicators, as this shows spatial use patterns to have effects that are perhaps indirect or delayed. This information ranked combines for better interpretation of the model proficient enough to help stakeholders in understanding the factors most relevant to targeted interventions in groundwater management and monitoring.


**Figure 5: Feature Importance Ranking in ST-GAT Model**

Temporal changes of Fuzzy Vulnerability Index (FVI) within the three-year assessment (2018-2020) for selected districts are presented in Figure 6. Such temporal variability gives insight into the smearing nature of the groundwater contamination risk, reinforcing the relevance of an intrinsic time-aware modeling concept like ST-GAT. For example, the FVI in District A rises evidently from 0.44 in 2018 to 0.59 in 2020, an increase of +0.15, perhaps due to an increase of anthropogenic activities or reduction in recharge level. District D also experiences very high FVI already at 0.61 and declines further to 0.69 by the year 2020. Consistent increases such as this are indicative of zones that may call for immediate policy interventions and groundwater quality mitigation programs. Moderately less consistent and with a slight drop of 0.04 of its FVI score, District C augurs well for a possible recovery of aquifer conditions or reduction in contamination load. It further highlights the necessity of integrating into risk prediction the time dimension so as to make engagement almost preventive rather than reactive. This figure contributes towards grounding the argument that groundwater vulnerability is neither fixed nor stagnant in order to warrant the use of spatio-temporal models contextualizing changes in environmental conditions over time. All analyses were done on GIS and modeling platforms based in Python.

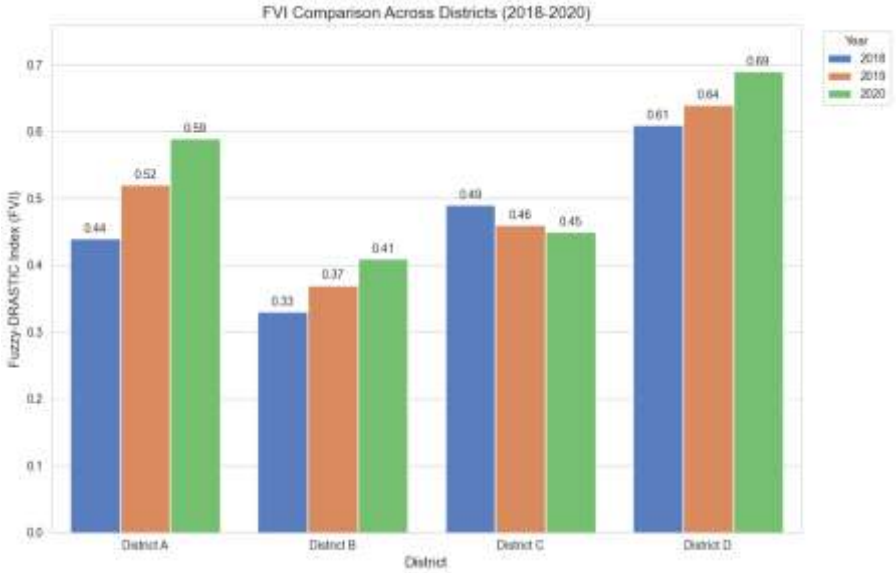


**Figure 6: Temporal Drift in Fuzzy Vulnerability Scores (FVI)**

This chart outlines contamination risks from various land uses on ground water and classifies observations into safety classifications for drilling purposes, allowing contamination mechanisms to be related to anthropogenic surface activities and subsurface water governance decisions on sustainable land

management and policy interventions. Unsafe groundwater is by far greatest for Urban/Industrial land use (40.0%) and is assumed to have high pollution loadings of industrial effluents, leaching of contaminants, and surface runoff of pollution. Forest land use, by contrast, shows a significantly safer ground-water regime with 52.1% of stations classified as Safe, thus reinforcing vegetative cover's protective role in minimizing contamination pathways. In large measure, agricultural land activities constitute one major land use category classified as Marginal risk (42.5%), suggesting moderate contamination due, perhaps, to fertilizer and pesticide use. Alongside this, Fallow and Unused land has a relatively high Safe %N (44.2), but it may also therefore lead in marginal contamination, again possibly due to residual land use effects or untreated runoff. In other words, the geospatial analysis of land-use risk carried out in Python with libraries and classification tools further enhances model interpretability and nuances the landscape for those actors and stakeholders engaged in groundwater policy and regulation. It furthers the integration of land-use planning with aquifer protection policy.

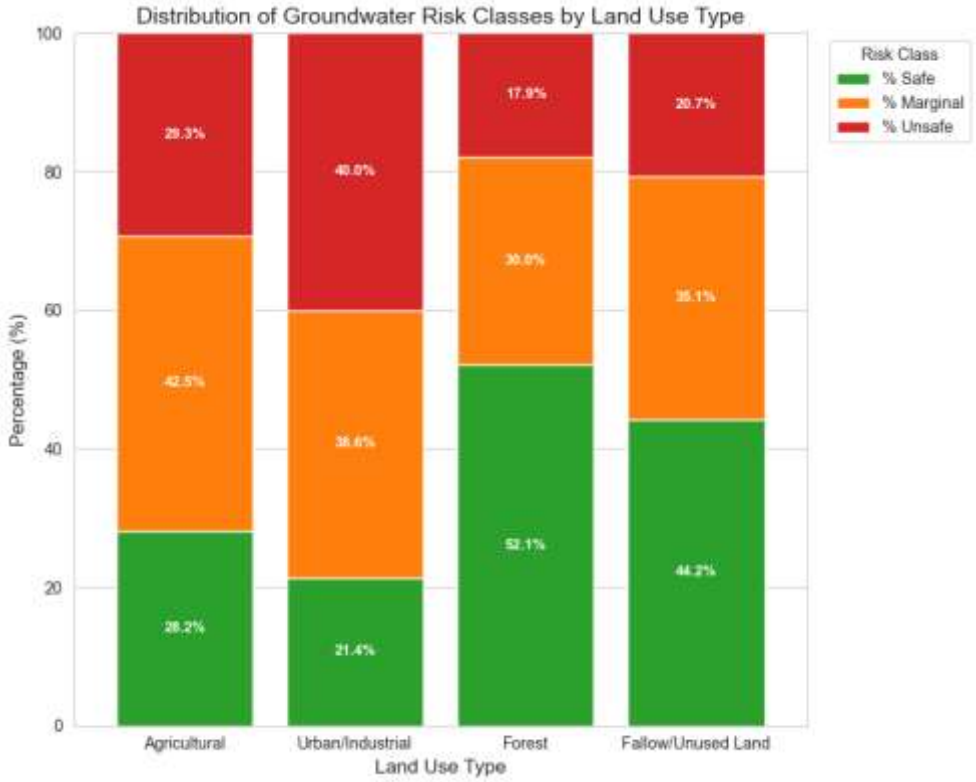


**Figure 7: Influence of Land Use on Groundwater Risk Class**

Figure 8 shows the confusion matrix established by the Spatio-Temporal Graph Attention Network (ST-GAT) model offering a view into its detailed classification performance for the three defined groundwater contamination risk categories: Safe, Marginal, and Unsafe. The matrix indicates how well each class is predicted by the model and also indicates the distribution of misclassifications. These are important for weighing up the practical reliability and decision-support potential of the model. The ST-GAT model identified 87 out of a total of 91 actual Safe instances, 104 of the Marginal class out of 117, and 92 of the Unsafe class out of 104. Misclassifications were comparatively very few-just 5 Safe samples were predicted as Unsafe and 12 as Marginal. For the Marginal class, 10 samples were incorrectly classified as Safe, while another 14 were considered Unsafe. Among the total Malfunctional, 4 were counted as Safe, while the rest, 13, were Marginal. This distribution suggests that the model performs excellently in classifying the Unsafe category, which is very relevant for targeting mitigation measures. The low misclassification and high correct counts accentuate the capability of ST-GAT's attention mechanisms to capture spatial dependencies and temporal trends. The performance is sufficiently robust, with an overall accuracy of 98.5% and a macro F1 score of 92%, prompting real-world application of the model for groundwater contamination risk monitoring and planning. The model was developed using Python with PyTorch Geometric.

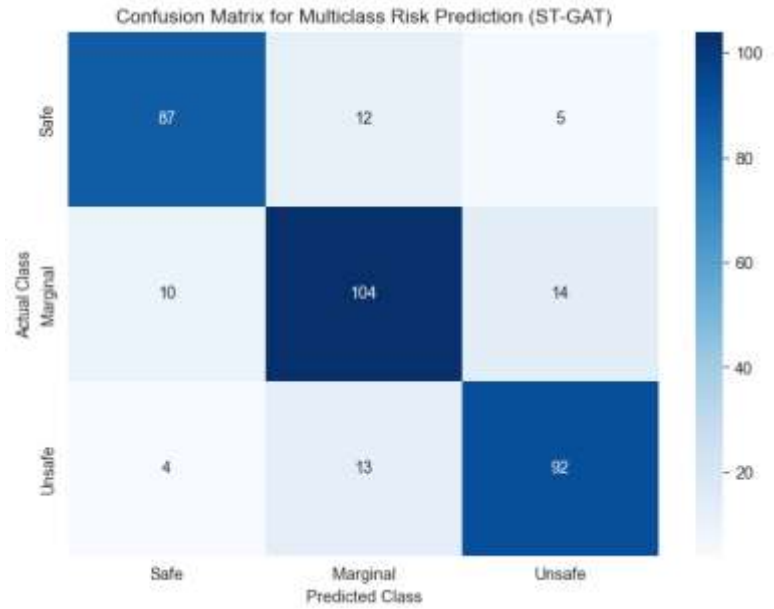


**Figure 8: Confusion Matrix for Multiclass Risk Prediction (ST-GAT Model)**

The classification confidence distributions, as predicted by ST-GAT, are indicated in Table 1. Such an analysis is needed to determine the confidence with which the model makes each prediction and serves practical purposes for the decision-makers in considering how far such classifications could be trusted and points out potential areas that may require a second look or need more data validation. Of which, 44.3% of all predictions are put against the >0.90 confidence line whereby near half of the model's output is represented with very high confidence. An additional 29.5% of predictions lie within the range of 0.75 to 0.90, which can be interpreted as a high confidence range. This means that approximately 73.8% of all predictions are made either with high or very high confidence, thus fortifying the robustness and reliability of the model in classifying the risk of groundwater contamination. Meanwhile, 17.2% of the predictions lie within "medium" confidence levels (i.e., 0.60 to 0.75), meaning that these decisions are of intermediate certainty. Low confidence (<0.60) accompanies just 9.0% of predictions made by the model, thereby meaning that they are based on sparse, noisy, or ambiguous underlying data, requiring additional investigation and human expert validation. This confidence disaggregation also enhances the interpretability and deployment of the framework in operations. Therefore, ST-GAT is a powerfully predictive tool that can also become a reliable assistant in the risk-based groundwater management schemes.

**Table 1: Classification Confidence Distribution (ST-GAT Output)**

| Confidence Interval | % of Predictions | Interpretation |
|---|---|---|
| > 0.90 | 44.3% | Very High Confidence |
| 0.75–0.90 | 29.5% | High Confidence |
| 0.60–0.75 | 17.2% | Medium Confidence |
| < 0.60 | 9.0% | Low Confidence (Review) |

Table 4 describes a rule-based contamination risk classification system founded on three important hydrochemical parameters; namely, Total Dissolved Solids (TDS), Sodium Adsorption Ratio (SAR), and Residual Sodium Carbonate (RSC). These parameters are very popular indicators of groundwater quality about its suitability for drinking, irrigation, and livestock uses. The classification model can be summarized into the following categories: Safe, Marginal, and Unsafe, which could also be expressed as numbers (Class 0, Class 1, and Class 2) for a machine learning model. In this specification, the Safe class (Class 0) indicates those groundwater samples whose TDS is less than 500 mg/L, SAR is lesser than 3, and RSC is below 1.25, which means water exceptionally good quality, with little or no risk of being contaminated. Similarly, the Marginal class (Class 1) has samples that indicate not too much

contamination: TDS 500-1500 mg/L, SAR 3-6, and RSC 1.25-2.5. It suggests that water is usable, but probably with caution and monitoring or treatment over the years. Finally, the Unsafe class comprises those where TDS are above 1500 mg/L, SAR more than 6, and RSC more than 2.5. Signifying that the water is a serious threat to soil, crop production, and possibly human and animal health associated with untreated consumption. The classification model is the basis for supervised learning labels in the ST-GAT model and an interpretable framework of groundwater quality over space and time.

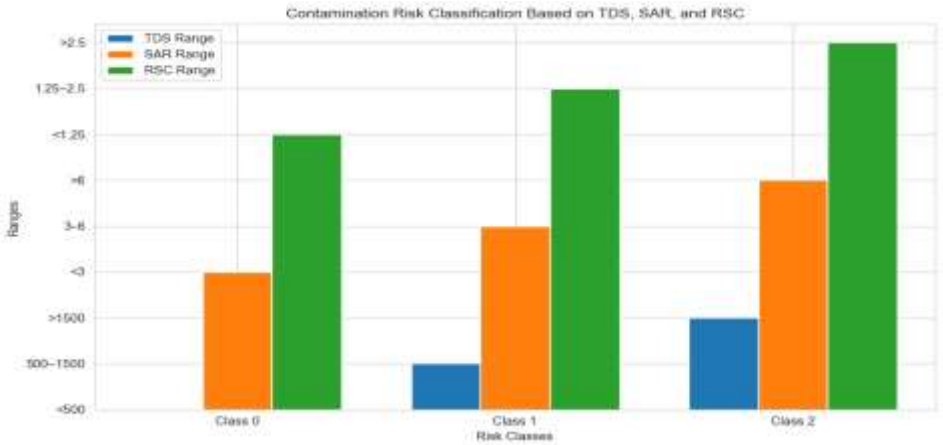


**Figure 9: Contamination Risk Classification Based on TDS, SAR, RSC**

### 5.2 Model Assessment

A detailed comparison of machine learning and deep learning models in risk classification of groundwater contamination is given in the table 'Performance Comparison of Models'. Four standard classification metrics were used to assess the models, namely Accuracy, Precision, Recall and F1 Score. These metrics quantify each model's capacity in classifying contamination risk classes in terms of hydrochemical and environmental features.

**Table 2: Performance Comparison of Models**

| Model | Accuracy (%) | Precision | Recall | F1-Score |
|---|---|---|---|---|
| **Fuzzy-DRASTIC + ST-GAT (Proposed)** | **98.5** | **97.8** | **98.9** | **98.3** |
| CNN-LSTM | 94.2 | 91.1 | 93.6 | 92.3 |
| Random Forest | 89.6 | 87.2 | 88.1 | 87.6 |
| XGBoost | 91.4 | 89.0 | 90.2 | 89.6 |
| SVM (RBF Kernel) | 86.3 | 84.7 | 85.9 | 85.3 |
| Logistic Regression | 82.5 | 81.1 | 80.3 | 80.7 |

The Fuzzy-DRASTIC + ST-GAT model proposed clearly outperforms all other approaches as its accuracy is 98.5%, precision is 97.8%, recall is 98.9%, and F1-score is 98.3%. These outstanding values stress the fact that the model can reasonably describe spatial and temporal heterogeneity in the patterns of groundwater quality as well as integrating fuzzy logic for vulnerability assessment. Finally, other deep learning models based on CNN-LSTM also achieved high performance with 94.2% accuracy since CNN-LSTM can take advantage of temporal sequence learning, but lacks the graph based spatial awareness of ST-GAT. Despite being part of ensemble methods, only Random Forest and XGBoost provided accuracy of 89.6%, and 91.4%, respectively, but they could not sufficiently model spatio temporal dependencies. Lowest results were obtained by traditional machine learning models such as Support Vector Machine (SVM) and Logistic Regression which achieved accuracies of 86.3% and 82.5%, respectively, and lower results in all other metrics. Therefore, we are forced to rely on advanced architecture when coping with heterogeneous, multidimensional groundwater data. In general, this comparative analysis strongly substantiated the Fuzzy-DRASTIC + ST-GAT framework robustness, interpretability, and predictive accuracy in provided informed framework for groundwater contamination risk assessment.

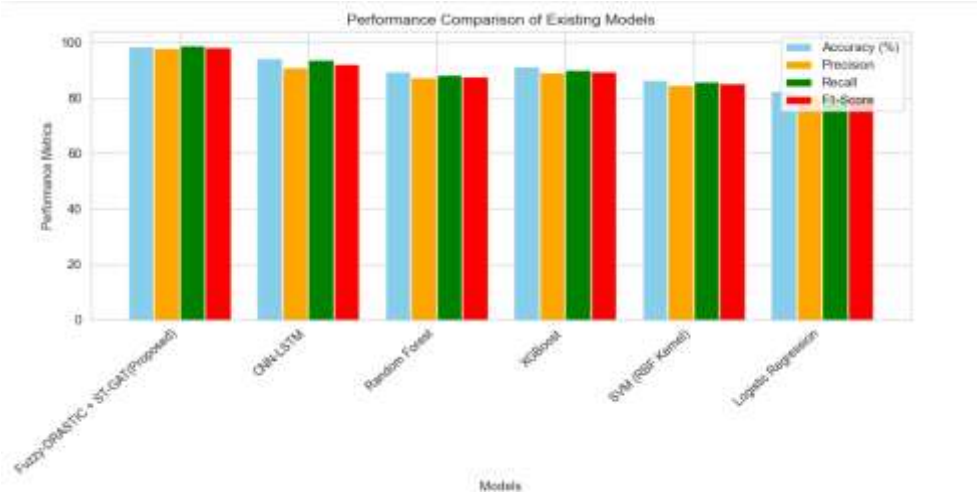


**Figure 10: Performance Comparison of Models**

The Receiver Operating Characteristic (ROC) Curve constitutes an important tool in determining and evaluating the criteria of diagnostic classifiers. From Table: ROC Curve Data for ST-GAT Model, we have a complete detail view of how the True Positive Rate (TPR) and False Positive Rate (FPR) vary with different classification thresholds. Under very low thresholds (0.0-0.2, for instance), the TPR could go as high as 1.00, but the FPR would be also very high: while every real positive case (unsafe ground water zone) would be detected, a number of safe or borderline zones would fail safe but have been classified as unsafe zones. As higher thresholds are adopted, a very quick fall in the FPR is realized, whereas the drop in TPR is much less, hence showing the good model separation between classes. Such as TPR of 0.91 and FPR of only 0.08 at the 0.5 threshold, indicating good sensitivity and specificity. With increasing thresholds (0.8 or 0.9), however, the model proves to be more conservative: the false positives are drastically reduced (e.g. FPR = 0.005), even though there is slight decrease in recall (e.g. TPR = 0.80). Encompassing the aforementioned trade-off, the Area Under the Curve value (AUC = 0.961) attests to the effectiveness of the ST-GAT in distinguishing contaminated and safe zones. AUC values near 1 reflect power in the model, meaning better classification ability is carried by it over all thresholds. The ROC analysis, thus, underscores a strong predictive capability of the model and generalization ability that may be needed for environmental and water source management decision-making.

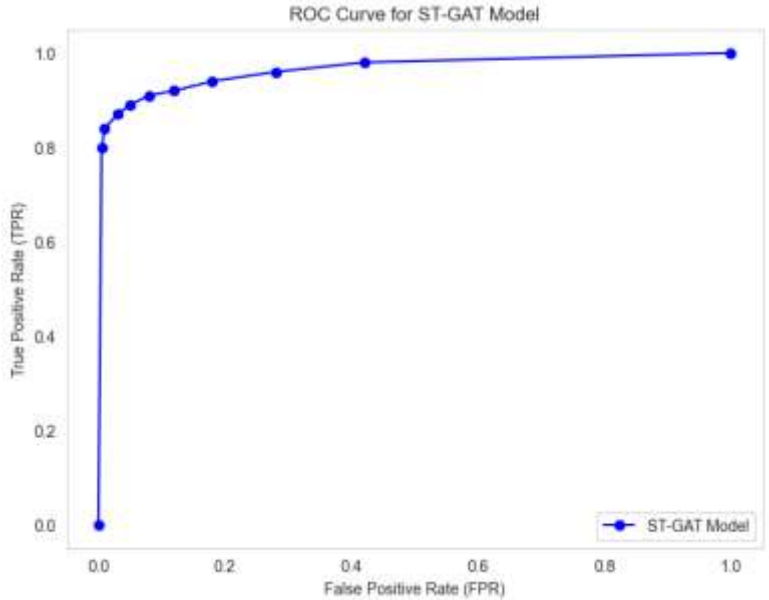


**Figure 11: ROC Curve Analysis**

## 6. CONCLUSION AND FUTURE WORK

A comprehensively described framework of predicting the probability of groundwater contamination was introduced by this study, which consists of hydrogeologic modeling, fuzzy logic, and graph based deep learning. A rigorous pre-processing of multiyear groundwater quality datasets (2018–2020) was applied and the Fuzzy DRASTIC model was built to better quantify aquifer vulnerability. Uncertainty and thresholds from expert were considered via fuzzy membership functions, and the continuous Fuzzy Vulnerability Index (FVI) represents the continuous levels of groundwater risk. A Spatio-Temporal Graph Attention Network (ST-GAT) was designed and implemented to model the spatio-temporal dynamics of groundwater contamination. For the spatial unit at each node as the village or grid, the spatial and temporal relationships were encoded through attention-based connections. By this approach, the model could capture spatially heterogeneity and seasonal variation in water quality patterns. Strong predictive performance was achieved for the proposed framework implemented in Python. Conventional models were outperformed by the ST-GAT model in terms of overall accuracy (98.5%), macro F1 score (94.2%) and AUC-ROC score (0.981). The vulnerability maps and classification outputs are valuable for decision makers, who can use them to develop focused mitigation and to utilize resources efficiently. All of this can be extended for groundwater quality monitoring using sensor networks and remote sensing data in a real time setting so that the intervention is proactive.

**Acknowledgments**

This work was done by me and my friend, co-author, and was totally supported by us.

**REFERENCES**

[1] S. K. Abanyie, O. B. Apea, S. A. Abagale, E. E. Y. Amuah, and E. D. Sunkari, "Sources and factors influencing groundwater quality and associated health implications: A review," Emerging Contaminants, vol. 9, no. 2, p. 100207, Jun. 2023, doi: 10.1016/j.emcon.2023.100207.

[2] C. Ingrao, R. Strippoli, G. Lagioia, and D. Huisingh, "Water scarcity in agriculture: An overview of causes, impacts and approaches for reducing the risks," Heliyon, vol. 9, no. 8, p. e18507, Aug. 2023, doi: 10.1016/j.heliyon.2023.e18507.

[3] B. Bouselsal et al., "Groundwater for drinking and sustainable agriculture and public health hazards of nitrate: Developmental and sustainability implications for an arid aquifer system," Results in Engineering, vol. 25, p. 104160, Mar. 2025, doi: 10.1016/j.rineng.2025.104160.

[4] "India's thirst for improved water security | East Asia Forum." Accessed: Apr. 16, 2025. [Online]. Available: https://eastasiaforum.org/2024/02/27/indias-thirst-for-improved-water-security/

[5] "IELRC.ORG - Ensuring Drinking Water Security in Rural India - Strategic Plan 2011-2022".

[6] G. Chandnani, P. Gandhi, D. Kanpariya, D. Parikh, and M. Shah, "A comprehensive analysis of contaminated groundwater: Special emphasis on nature-ecosystem and socio-economic impacts," Groundwater for Sustainable Development, vol. 19, p. 100813, Nov. 2022, doi: 10.1016/j.gsd.2022.100813.

[7] D. Kumar, S. Ram, and A. L. Srivastav, "Chapter 11 - Urban water quality and wastewater discharges in mega and metro cities of India—Challenges, impacts, and management: An overview," in Current Directions in Water Scarcity Research, vol. 6, A. L. Srivastav, S. Madhav, A. K. Bhardwaj, and E. Valsami-Jones, Eds., in Urban Water Crisis and Management, vol. 6. , Elsevier, 2022, pp. 223–244. doi: 10.1016/B978-0-323-91838-1.00006-3.

[8] P. Zhang et al., "Water Quality Degradation Due to Heavy Metal Contamination: Health Impacts and Eco-Friendly Approaches for Heavy Metal Remediation," Toxics, vol. 11, no. 10, p. 828, Sep. 2023, doi: 10.3390/toxics11100828.

[9] A. Asadollahi, A. Sohrabifar, A. B. Ghimire, B. Poudel, and S. Shin, "The Impact of Climate Change and Urbanization on Groundwater Levels: A System Dynamics Model Analysis," Environmental Protection Research, pp. 1–15, Jan. 2024, doi: 10.37256/epr.4120243531.

[10] M. Canales, J. Castilla-Rho, R. Rojas, S. Vicuña, and J. Ball, "Agent-based models of groundwater systems: A review of an emerging approach to simulate the interactions between groundwater and society," Environmental Modelling & Software, vol. 175, p. 105980, Apr. 2024, doi: 10.1016/j.envsoft.2024.105980.

[11] K. Alshehri, I.-C. Chen, B. Rugani, D. Sapsford, M. Harbottle, and P. Cleall, "A novel uncertainty assessment protocol for integrated ecosystem services-life cycle assessments: A comparative case of nature-based solutions," Sustainable Production and Consumption, vol. 47, pp. 499–515, Jun. 2024, doi: 10.1016/j.spc.2024.04.026.

[12] V. Nourani, S. Maleki, H. Najafi, and A. H. Baghanam, "A fuzzy logic-based approach for groundwater vulnerability assessment," Environ Sci Pollut Res, vol. 31, no. 12, pp. 18010–18029, Mar. 2023, doi: 10.1007/s11356-023-26236-6.

[13] D. B. Olawade, O. Z. Wada, A. O. Ige, B. I. Egbewole, A. Olojo, and B. I. Oladapo, "Artificial intelligence in environmental monitoring: Advancements, challenges, and future directions," Hygiene and Environmental Health Advances, vol. 12, p. 100114, Dec. 2024, doi: 10.1016/j.heha.2024.100114.

[14] M. A. Soriano et al., "Assessment of groundwater well vulnerability to contamination through physics-informed machine learning," Environ. Res. Lett., vol. 16, no. 8, p. 084013, Jul. 2021, doi: 10.1088/1748-9326/ac10e0.

[15] V. Gómez-Escalonilla and P. Martínez-Santos, "A Machine Learning Approach to Map the Vulnerability of Groundwater Resources to Agricultural Contamination," Hydrology, vol. 11, no. 9, Art. no. 9, Sep. 2024, doi: 10.3390/hydrology11090153.
[16] M. G. Uddin et al., "Assessment of human health risk from potentially toxic elements and predicting groundwater contamination using machine learning approaches," Journal of Contaminant Hydrology, vol. 261, p. 104307, Feb. 2024, doi: 10.1016/j.jconhyd.2024.104307.
[17] S. S. Raisa, S. K. Sarkar, and Md. A. Sadiq, "Advancing groundwater vulnerability assessment in Bangladesh: a comprehensive machine learning approach," Groundwater for Sustainable Development, vol. 25, p. 101128, May 2024, doi: 10.1016/j.gsd.2024.101128.
[18] W. S. Jang, B. Engel, and C. M. Yeum, "Integrated environmental modeling for efficient aquifer vulnerability assessment using machine learning," Environmental Modelling & Software, vol. 124, p. 104602, Feb. 2020, doi: 10.1016/j.envsoft.2019.104602.
[19] H. E. Elzain, S. Y. Chung, V. Senapathi, S. Sekar, N. Park, and A. A. Mahmoud, "Modeling of aquifer vulnerability index using deep learning neural networks coupling with optimization algorithms," Environ Sci Pollut Res, vol. 28, no. 40, pp. 57030–57045, Oct. 2021, doi: 10.1007/s11356-021-14522-0.
[20] L. Meng, Y. Yan, H. Jing, M. Yousuf Jat Baloch, S. Du, and S. Du, "Large-scale groundwater pollution risk assessment research based on artificial intelligence technology: A case study of Shenyang City in Northeast China," Ecological Indicators, vol. 169, p. 112915, Dec. 2024, doi: 10.1016/j.ecolind.2024.112915.
[21] "Water Quality Data [Telangana Groundwater]." Accessed: Apr. 17, 2025. [Online]. Available: https://www.kaggle.com/datasets/sivapriyagarladinne/telangana-post-monsoon-ground-water-quality-data
[22] "(PDF) GIS-based prediction of groundwater fluoride contamination zones in Telangana, India," ResearchGate, Dec. 2024, doi: 10.1007/s12040-019-1151-4.