



AutoML-GWL: Automated machine learning model for the prediction of groundwater level

Abhilash Singh ^{a,*}, Sharad Patel ^{a,1}, Vipul Bhadani ^{a,1}, Vaibhav Kumar ^b, Kumar Gaurav ^{a,*}

^a Fluvial Geomorphology and Remote Sensing Laboratory, Department of Earth and Environmental Sciences, Indian Institute of Science Education and Research Bhopal, Bypass road, Bhopal, 462066, Madhya Pradesh, India

^b Geospatial Artificial Intelligence (GeoAI) Laboratory, Department of Data Science and Engineering, Indian Institute of Science Education and Research Bhopal, Bypass road, Bhopal, 462066, Madhya Pradesh, India

ARTICLE INFO

Keywords:

AutoML
Groundwater
Bayesian optimisation
Machine learning

ABSTRACT

Predicting groundwater levels is pivotal in curbing overexploitation and ensuring effective water resource governance. However, groundwater level prediction is intricate, driven by dynamic nonlinear factors. To comprehend the dynamic interaction among these drivers, leveraging machine learning models can provide valuable insights. The drastic increase in computational capabilities has catalysed a substantial surge in the utilisation of machine learning-based solutions for effective groundwater management. The performance of these models highly depends on the selection of hyperparameters. The optimisation of hyperparameters is a complex process that often requires application-specific expertise for a skillful prediction. To mitigate the challenge posed by hyperparameter tuning's problem-specific nature, we present an innovative approach by introducing the automated machine learning (AutoML-GWL) framework. This framework is specifically designed for precise groundwater level mapping. It seamlessly integrates the selection of best machine learning model and adeptly fine-tunes its hyperparameters by using Bayesian optimisation. We used long time series (1997-2018) data of precipitation, temperature, evaporation, soil type, relative humidity, and lag of groundwater level as input features to train the AutoML-GWL model while considering the influence of Land Use Land Cover (LULC) as a contextual factor. Among these input features, the lag of groundwater level emerged as the most relevant input feature. Once the model is trained, it performs well over the unseen data with a strong correlation of coefficient ($R = 0.90$), low root mean square error ($RMSE = 1.22$), and minimal bias = 0.23. Further, we compared the performance of the proposed AutoML-GWL with sixteen benchmark algorithms comprising baseline and novel algorithms. The AutoML-GWL outperforms all the benchmark algorithms. Furthermore, the proposed algorithm ranked first in Friedman's statistical test, confirming its reliability. Moreover, we conducted a spatial distribution and uncertainty analysis for the proposed algorithm. The outcomes of this analysis affirmed that the AutoML-GWL can effectively manage data with spatial variations and demonstrates remarkable stability when faced with small uncertainties in the input parameters. This study holds significant promise in revolutionising groundwater management practices by establishing an automated framework for simulating groundwater levels for sustainable water resource management.

1. Introduction

Globally, groundwater constitutes nearly 97% of the Earth's available freshwater resources (Ravenscroft and Lytton, 2022). This essential resource plays a pivotal role in various facets of human existence, encompassing critical domains like drinking water provision, agricultural sustenance, and industrial advancement (Tao et al., 2022; Niranjanaik

et al., 2022). The significance of groundwater is underscored by the alarming trends observed in major aquifers worldwide. Startlingly, 21 out of the world's 34 largest aquifers are currently grappling with swift and unsustainable depletion, a concerning situation highlighted by organisations like the United Nations. Evidently, countries such as India, China, and the United States find themselves at the epicentre

* Corresponding authors.

E-mail addresses: sabhilash@iiserb.ac.in, abhilash.iiserb@gmail.com (A. Singh), sharadp@iiserb.ac.in (S. Patel), vipul18@iiserb.ac.in (V. Bhadani), vaibhav@iiserb.ac.in (V. Kumar), kgaurav@iiserb.ac.in (K. Gaurav).

¹ Authors contributed equally.

² <https://ecology.wa.gov/>.

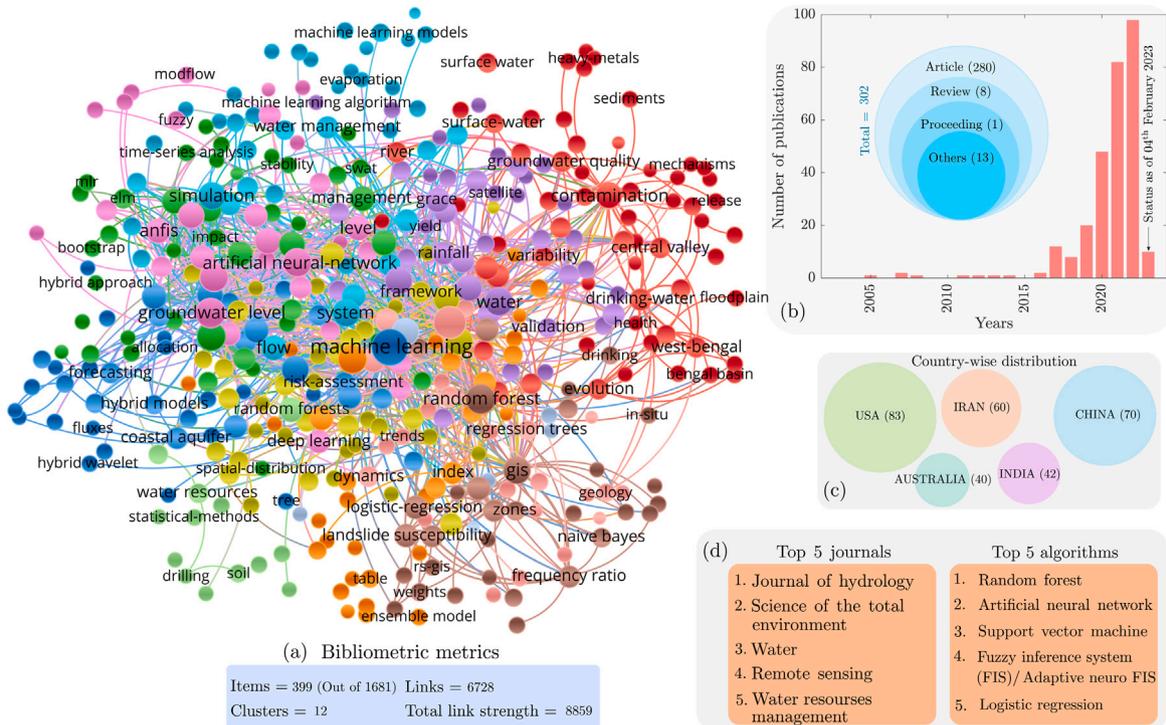


Fig. 1. Bibliometric analysis (a) shows the keywords (*i.e.*, items) that have appeared at least twice in research publications (indexed in Web of Science) that have been published in the last 20 years concerning groundwater level prediction using machine learning, (b) bar graph showing the year-wise publication along with the document types, (c) shows the country-wise distribution, and (d) shows the top 5 venues where the research has been published along with the top five machine learning algorithms that have been used extensively in groundwater level prediction.

of this crisis (Zhongming et al., 2021). The outcomes of this depletion are significant and far-reaching, encompassing a heightened likelihood of floods, prolonged droughts, and expanding desert areas. Moreover, there is a noticeable decrease in agricultural productivity, affecting both crop yields and livestock production. Beyond the economic and public health implications that arise from this depletion, it also brings about grave environmental challenges. These encompass the destruction of habitats, a decline in biodiversity, and disruptions to the delicate equilibrium of natural systems. As a concrete example, consider the diminishing irrigated wetlands in regions like the United States, which places rare plant species, birds, and fish at risk of extinction. Hence, it is of utmost importance to protect and monitor the global groundwater reserves for the benefit of current and future generations.² It is imperative to implement measures that facilitate the maintenance and restoration of healthy GWL while simultaneously curbing contamination and pollution within our groundwater sources. This concerted effort is essential to ensure a consistent and dependable water supply for the inhabitants of our planet. The pivotal role of GWL extends to effective land use planning and informed decision-making (Lerner and Harris, 2009; Chambel, 2015). The continuous monitoring of the GWL helps to understand the spatiotemporal dynamics of the subsurface water resources and its seasonal variations (Patel and Rastogi, 2017). This vigilant monitoring contributes significantly to our comprehension of groundwater's behaviour, enhancing our ability to accurately forecast the repercussions of climate change and other environmental variables (Azizi et al., 2021). Moreover, an up-to-date evaluation of GWL holds the key to effective water resource management and urban planning. By providing invaluable insights, it aids authorities in making informed choices that reverberate across various sectors. On the whole, groundwater level measurements play an essential role in understanding the water cycle and its impact on the environment.

Diverse numerical techniques offer avenues to estimate GWL, each contributing to a multifaceted understanding. These encompass analytical methods, such as comparative water level indices and geostatistical

strategies, including kriging and co-kriging (Ahmadi and Sedghamiz, 2008). Analytical techniques are commonly used to compare water level indices across time or to determine the depth of the water table. These techniques are also used to collect and analyse data from multiple sources. Apart from that, the numerical methods-based simulation models (groundwater flow models) are computer-based modelling tools that are used to determine groundwater levels and movement. Numerical methods like finite difference method (FDM; Omar et al. (2019)), finite element method (FEM; Narasimhan and Witherspoon (1982)) and recently developed meshfree method (Patel et al., 2022) are used to simulate and predict the spatio-temporal changes in water table by simulating the physical systems and processes. These methods are employed to predict the changes in water levels due to pumping and injection activities in the surrounding environment. These techniques can provide detailed information about water tables and can help identify areas where groundwater resources are abundant or limited. Although these numerical techniques are successfully applied to a variety of aquifer problems, however, there are several limitations as well. Like, analytical techniques are limited in their accuracy due to several factors such as the availability and quality of data, uncertain boundary conditions, and the need for large amounts of data (Meza-Gastelum et al., 2022). Geostatistical techniques like kriging and co-kriging can be subject to errors resulting from data interpolation, spatial trends, and temporal variations (Ahmadi and Sedghamiz, 2007). Furthermore, while numerical models offer valuable insights, they demand sophisticated computer programs and highly detailed data, presenting challenges for practical implementation in the field. It is worth noting that these methods can sometimes yield less accurate results due to the simplifying assumptions inherent in their implementation (Szidarovszky et al., 2007). In parallel, the realm of groundwater assessment can also benefit from the utilisation of remote sensing imagery, providing additional layers of information about surface water features (Coelho et al., 2017). However, it is important to consider that the effectiveness of employing remote sensing

images to map surface water features hinges on appropriate imagery selection and suitable resolution; otherwise, inaccurate predictions may arise (Ahamed et al., 2022).

The limitations inherent in numerical modelling and remote sensing techniques can be effectively addressed by embracing the machine learning approach (Tao et al., 2022). In recent years, the application of machine learning has gained considerable momentum in predicting various facets of groundwater dynamics. This versatile approach has been leveraged to forecast an array of critical factors, including sediment discharge, surface ozone gas concentration, water quality index, river flow forecasting, heavy sediment metal levels, Standardised Precipitation Evapotranspiration Index (SPEI), and drought occurrences. For instance, a noteworthy case study conducted on the Narmada River in India employed the backpropagation neural network training algorithm to accurately predict daily sediment discharge (Bisoyi et al., 2019). Similarly, a robust artificial intelligence framework facilitated the multi-hour advance prediction of surface ozone gas concentration (AlOmar et al., 2020). In tropical regions, such as Malaysia, machine learning played a pivotal role in predicting the Water Quality Index (WQI) (Hameed et al., 2017). Moreover, the deployment of advanced techniques like Radial Basis Function Neural Networks (RBFNN) and Feed Forward Neural Networks (FFNN) enabled precise forecasting of daily river flow levels in the Johor River, Malaysia (Yaseen et al., 2016). Notably, hybrid Artificial Intelligence (AI) models were harnessed to predict sediment-heavy metal levels across Australian bays (Bhagat et al., 2021). Another significant advancement came in the form of a FFNN-based model developed to predict monthly SPEI using hydrometeorological parameters and climate indices in Eastern Australia (Deo and Şahin, 2015). Lastly, Artificial Neural Networks (ANN) were used to predict droughts in dry lands (Buckland et al., 2019). In essence, the aforementioned studies vividly underscore the substantial impact of machine learning in constructing accurate and dependable models for predicting diverse aspects of groundwater dynamics.

In this study, we mainly focused on machine learning approach for groundwater level prediction, as the knowledge of GWL variations can be used to quantify groundwater availability. Fig. 1a shows the result of bibliometric analysis of the words groundwater level and machine learning. The number of studies concerning groundwater level and machine learning has increased exponentially in the last 20 years. A total of 302 research items were published (in Web Of Science), which includes articles, reviews, and conference proceedings with USA and China as the leading countries (Fig. 1b&c). We extracted the author keywords from these publications that have appeared in at least two research publications. We found that of 1681 keywords, 399 have appeared at least twice. All these 399 keywords (or items) were clustered in 12 clusters based on the VOS clustering algorithm (Van Eck and Waltman, 2011) and are shown in 12 different colours (Fig. 1a). A link between two items or keywords (within the same cluster or with different clusters) shows the bibliographic coupling between them. The link strength indicates the number of publications in which these two words have appeared together, and the total link strength is the sum of all the link strength (Singh et al., 2023b). Based on the total link strength, we have sorted some critical insights such as publications venue and frequently used algorithms for groundwater level mapping (Fig. 1d).

Numerous endeavors have been undertaken to map fluctuations in groundwater levels by harnessing machine learning in conjunction with meteorological variables. For instance, Wen et al. (2015) employed a wavelet-based adaptive neuro-fuzzy model to map GWL from GWL lag. This endeavor, centred on two wells located in Laizhou, China, covered a two-year span from 2007 to 2009. The study reported an impressive predictive capacity, with an R value ranging from 0.75 to 0.98. However, the limitations encompassed the narrow data scope and the limited number of wells. Another investigation by Nair and Sindhu (2016) utilised ANN to predict GWL, drawing on data from rainfall,

evapotranspiration, temperature, and humidity. Spanning the period from 2002 to 2016, the study focused on the Mamon River basin in Kerala, India. The architecture, comprising a single layer of 5 to 15 neurons, exhibited substantial accuracy with R values ranging from 0.80 to 0.91. Despite the achievement, the constraint here was the small number of wells (only eleven). Nadiri et al. (2019) proposed a committee fuzzy logic model to map GWL in the Duzdüzam basin, Iran. This model incorporated GWL lag, discharge, rainfall, and temperature data from 2007 to 2016. The research showcased an impressive predictive prowess, yielding R values from 0.83 to 0.97. However, like the previous studies, the limitation persisted due to the small number of wells (only eight). In a quest for advancement, Banadkooki et al. (2020) introduced a modified radial neural network infused with whale optimisation to fine-tune hyperparameters. The scope extended to encompass different lags of precipitation and temperature for mapping GWL from 2000 to 2012. The Yazd-Ardakan region in Iran was the focus, and the model garnered strong performance with R values from 0.89 to 0.96. More recently, Pham et al. (2022) devised a bagging random tree model for GWL prediction, leveraging mean temperature, relative humidity, and rainfall data from 1981 to 2017. The study centred on two wells in the northwestern part of Bangladesh. Impressively, their proposed model surpassed other machine learning techniques, exhibiting R values ranging from 0.60 to 0.96 and RMSE values from 0.3 to 1.8. The limitation, however, was rooted in the low number of wells (only two). In the vanguard of innovation, Chidepudi et al. (2023) embarked on developing predictive models for GWL using wavelet-based deep learning models such as long short-term memory, gated recurrent unit, and bidirectional LSTM. This undertaking drew from fifty years of GWL data (1970–2020) from three wells. The study reported compelling predictive accuracy, showcasing R values from 0.86 to 0.93 and RMSE values from 1.2 to 1.7. Across the spectrum of these studies, a recurring constraint involves the limited number of wells under consideration. Additionally, manual algorithm selection poses potential biases if not benchmarked against a reference algorithm. Further, hyperparameter optimisation across different algorithms typically lacks a uniform approach within manual setups.

To overcome the limitations mentioned above, we proposed an automated machine learning model to predict the groundwater level. The model automatically optimises the required set of hyperparameters iteratively and selects the best algorithm based on their performance on the objective function (i.e., minimum loss). To do so, we considered a semi-arid region (i.e., Betwa river basin) in Central India. We divided the entire area of interest into 126 grids (0.25° each) comprising 665 wells. We used precipitation, temperature, evaporation, soil type, relative humidity, and lag of groundwater level at each grid from 1997 to 2018 to develop the AutoML-GWL model. We assess the performance of the returned model on the unseen datasets considering R, RMSE, and bias as accuracy metrics. The main contributions of this paper are summarised as follows:

- This study proposes a novel automated machine learning model, which returns the best model with optimised hyperparameters to predict the GWL. This model has generalisation capabilities and may be tuned for any specific region.
- The proposed AutoML-GWL model is trained and evaluated on huge datasets of 665 wells (from 1997 to 2018) by considering precipitation, temperature, evaporation, soil type, relative humidity, and GWL lag as potential input features.
- This study evaluates feature importance and sensitivity to understand the impact of each contributing feature on GWL fluctuations by leveraging regression tree ensemble learning.
- This study incorporates the impact of LULC on the prediction of GWL.

The rest of the manuscript is organised into four distinct sections. In Section 2, we delve into the intricacies of the datasets, outline the process of feature engineering, and provide a detailed account

of the model development—specifically, the AutoML-GWL approach. Section 3 is dedicated to discussing the results we obtained. This encompasses the exploration of feature importance, feature sensitivity, model performance, error analysis, and residual analysis. In Section 4, our focus shifts to the presentation of our discussions. Here, we draw comparisons with established benchmarks, conduct uncertainty analysis, delve into spatial distribution analysis, and assess the impact of our proposed model. Lastly, in Section 5, we draw our conclusions based on the findings and insights garnered throughout the study.

2. Material and methods

This section first discusses the input datasets used to develop the AutoML-GWL model. We discussed the data collection, pre-processing, and generation part. Afterward, we discussed the feature importance and sensitivity evaluation, followed by the model development steps. The flowchart in Fig. 2 illustrates the complete methodology.

2.1. Datasets

In this study, we used groundwater level data as the response variable to train the proposed AutoML-GWL model along with various input features. We collected the groundwater level time-series data (from 1997–2018) for 665 dug wells over the area of interest from the Madhya Pradesh Water Resource Board (MPWRD) and the Central Ground Water Board (CGWB). They provide the measurements of the water table's depth below the earth's surface in meters (mbgl). The data is usually measured four times yearly (*i.e.*, January, May, August, and October). We categorised these data into pre-monsoon (May), monsoon (August), and post-monsoon (October and January) according to Indian meteorological classification to study the seasonal groundwater level fluctuation (Bansod et al., 2003; Karim et al., 2012; Raturi et al., 2022; Dhawan et al., 2023).

We integrated the impact of Land Use Land Cover (LULC) to determine the representative groundwater level for each grid point. To achieve this, we employed publicly accessible LULC data sourced from ESRI, Microsoft, and Impact Observatory.³ This dataset was constructed using Sentinel-2 satellite imagery and encompasses a comprehensive LULC map comprising ten distinct classes – namely, trees, grass, water, crops, flooded vegetation, built-up areas, bare ground, shrubbery, snow, and clouds – at a spatial resolution of 10 m (Karra et al., 2021). We extracted the corresponding LULC class at each grid point and subsequently identified wells that fell within the major land class. Wells situated in other LULC classes were excluded from consideration, as illustrated in Fig. 3. This decision was based on the rationale that the dominant land use in a specific area would significantly influence the groundwater level dynamics in that region. Our methodology prioritised the major land class within each grid, leading us to employ the most recent LULC data available, which was dated April 11, 2022. We based this decision on the assumption that the primary land class within each grid would not experience significant and rapid changes. By incorporating this approach, we ensure a holistic consideration of LULC dynamics and their influence on groundwater levels across the study area.

To map the complex dynamics of the groundwater fluctuations, we considered precipitation, temperature, evaporation, soil type, and relative humidity as the potential input features to train the AutoML-GWL model. We downloaded gridded ERA5 reanalysis datasets of precipitation, temperature, evaporation, soil type, and relative humidity from the Climate Data Store (CDS)⁴ at a spatial resolution of at 0.25° from 1997 to 2018.

Precipitation includes accumulated liquid and frozen water that falls on the earth's surface in the form of rain and snow. It considers both convective as well as large-scale precipitation. This parameter has a unit of depth in meters of water equivalent. It is the depth of water that we get after being uniformly distributed throughout the grid. Temperature (in Kelvin) indicates the air temperature at a height of 2 m above the ground surface. Evaporation represents the total amount of water that gets evaporated from the ground surface. It is expressed in meters of water equivalent (m.w.e). CDS classifies the soil into seven categories depending on the water holding capacity of the soil collected below 100 cm from the top soil surface (*i.e.*, root zone data). The seven categories are coarse (soil 1), Medium (soil 2), Medium fine (soil 3), fine (soil 4), very fine (soil 5), organic (soil 6), and tropical organic (soil 7). Over the area of interest, only soil 2, 3, and 4 are available. Relative humidity is the water vapour pressure as a proportion of the saturation point for air. It is expressed in percentage. In addition to these five features, we have generated a synthetic feature by taking the one-lag of the groundwater level (*i.e.*, GWL lag). To do so, we computed the serial correlation of GWL through the auto-correlation function. We found that one-lag is the highest correlated lag; hence we only considered the one-lag. A detailed statistical description of the input and output data can be found in Table 1.

2.2. Feature importance

In machine learning, it is a standard practice to check the predictive relevance of the input features before training the machine learning model. Through feature importance analysis, we can identify the most and the least relevant feature for the prediction of the response variable. To do so, we used the regression tree ensemble approach to boost hundred ($\zeta = 100$) decision trees (using a least square boosting algorithm) by keeping unity learning rate (*i.e.*, $\beta=1$). The algorithm identifies the weakness in the first decision tree while training it and subsequently generates the next decision tree (g_j) by overcoming the weakness of the previous tree (Zhou and Hooker, 2021; Singh et al., 2021a; Nagar et al., 2023; Singh et al., 2022b). In this iterative mode, the current model (T_j) is updated by overcoming the weakness of the previous model (T_{j-1}) (see Eq. (1)).

$$T_j = T_{j-1} + (\alpha \cdot \beta \cdot g_j) \quad (j = 1, 2, \dots, \zeta) \quad (1)$$

where α_j is the weight for each decision tree. Finally, at $j = \zeta$, the model turned into a strong learner (*i.e.*, T_ζ). We then computed the relative importance score of each feature by computing the total change in the node risk (ΔR) resulting from the feature split and normalising it with respect to the number of the total branches (R_{branch}) using Eq. (2).

$$\Delta R = \frac{R_p - (R_{c1} + R_{c2})}{R_{branch}} \quad (2)$$

where R_p represents parent node risk and R_{c1} & R_{c2} represents the children node risk. The individual node risk (R_i) is computed from the probability of the node (P_i) and the corresponding mean square error (MSE_i) (see Eq. (3)).

$$R_i = P_i \cdot MSE_i \quad (3)$$

We also computed the feature association matrix, a 6-by-6 matrix that measures the correlation between the input features. It is a quantifiable indicator that reflects the degree of resemblance between decision rules utilised to partition data instances. Within decision tree construction, this measure serves as a metric for evaluating the likeness between potential decision splits and the ultimate optimal split, which is determined during the tree's growth. The predictive measure of association between any two features, say l and m is calculated by using Eq. (4).

$$P_{asso} = \frac{\min(P_{c1}, P_{c2}) - (1 - P_{c1c1m} - P_{c2c2m})}{\min(P_{c1}, P_{c2})} \quad (4)$$

where P_{c1} and P_{c2} represent the proportion of observations for the two children nodes, $c1$ and $c2$, respectively.

³ Available at: <https://www.arcgis.com/>.

⁴ <https://cds.climate.copernicus.eu>.

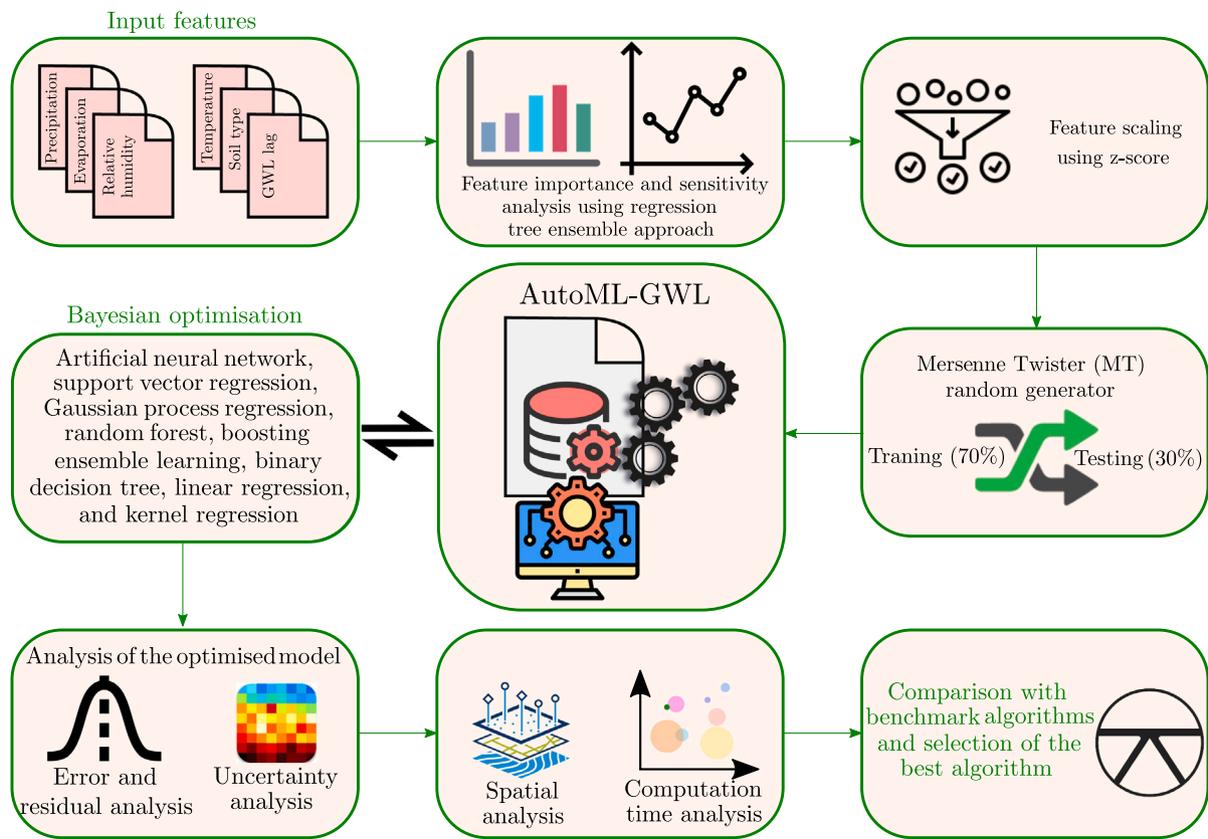


Fig. 2. Flowchart for the prediction of groundwater level using AutoML-GWL.

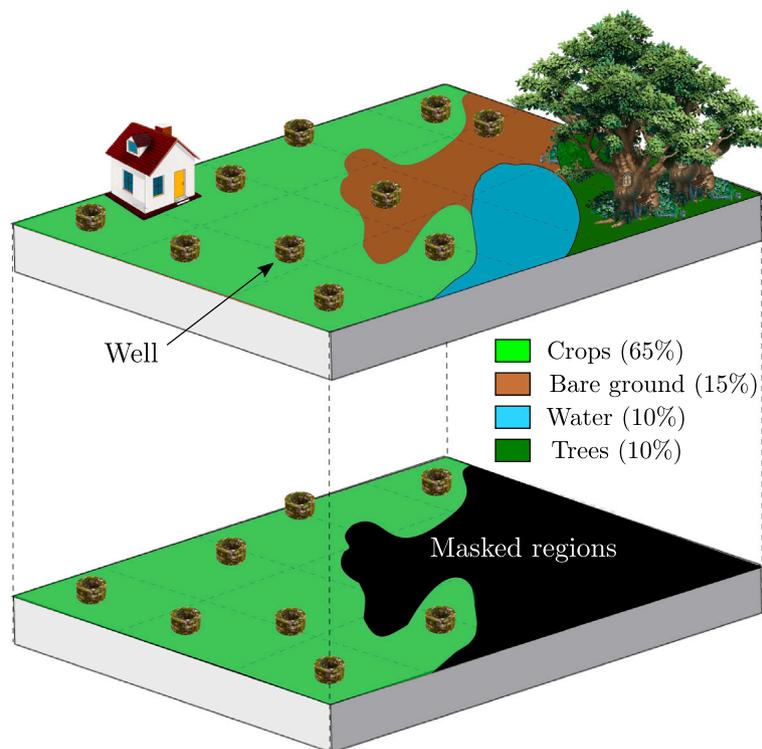


Fig. 3. Schematic illustration of the inclusion of LULC impact.

Table 1
Statistical description of the input and output data.

Datasets	Input features						Response variable
	Precipitation [m]	Temperature [K]	Evaporation [-m.w.e]	Soil type	Relative humidity [%]	GWL lag [m]	GWL [m]
Count	8310	8310	8310	8310	8310	8310	8310
Data type	Numerical	Numerical	Numerical	Categorical (Converted to Numerical)	Numerical	Numerical	Numerical
Maximum	0.1246	310.1780	0.0555	4	72.4992	35.8250	35.8250
Minimum	0	295.1570	0.0113	2	16.6241	0.0100	0.0100
Mean	0.0150	304.1833	0.0360	3.4606	40.5616	6.7167	6.6916
Median	0.0026	305.3469	0.0363	3.0000	36.9325	6.3500	6.3000
Range	0.1246	15.0209	0.0442	2	55.8751	35.8150	35.8150
Standard deviation	0.0203	3.3688	0.0083	0.5586	16.8748	3.8083	3.7971

2.3. Feature sensitivity

Feature importance analysis only tells us about the relative importance score rather than the impact (increasing, decreasing, or undulating) of each individual feature ($F = \{f_1, f_2, \dots, f_T\}$, where T is the total number of input features) on the response variable. To assess the impact of each feature, we computed the partial dependency plot (PDP), and individual condition expectation (ICE) curves (Friedman, 2001; Goldstein et al., 2015). Both these techniques are often used concurrently to study the impact of input features (Singh et al., 2021a; Singh and Gaurav, 2023; Singh et al., 2023a). ICE captures the impact of each feature at each discrete observation, whereas PDP only tells us about the overall impact by marginalising the impact of other features. The PDP of the T th feature is computed using Eq. (5).

$$f(f_T) \approx \frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} f(f_T, F - f_T) \quad (5)$$

where $f(\cdot)$ is the mapping function that was previously developed to compute feature importance. N_{obs} is the total number of observations. The ICE curves are computed by disaggregating Eq. (5). A detailed explanation of PDP and ICE curves for feature sensitivity can be found in Singh et al. (2023c).

After evaluating the feature sensitivity, we normalised the entire dataset using the standard z-score scaling technique (Evans, 2006). Following this, we utilised the Mersenne Twister (MT) random generator to split the complete data into a 70:30 ratio for training and testing the model (i.e., 70% for training and 30% for testing). The dimension (rows \times column) of the entire dataset is 8310×7 , where rows represent the number of observations. The first six columns represent the input features, and the last represents the response variable. The dimension of the training and testing datasets are 5817×7 and 2493×7 , respectively.

2.4. Automated machine learning

With the rapid increase in data-driven usage in solving real-life problems, the complexity concerning the model development part (i.e., feature engineering, model optimisation, etc.) increases drastically. The recently evolved automated machine learning (AutoML) model significantly eases the complications by automating the complete process. It eliminates the requirement of highly skilled expertise in applying machine learning to solve real-world research snags (Guyon et al., 2019; Liu et al., 2020; He et al., 2021). In recent years, it has shown its potential in industry as well as academics by solving complex real life problems (Singh et al., 2022a; Raj et al., 2023; Liuliakov et al., 2023; Sahin and Demir, 2023). Following this, we developed an automated machine learning model to predict the groundwater level, where we automated the model optimisation and selection process. A pseudocode of the proposed algorithm is shown in Algorithm 1. We have briefly discussed the participating algorithms in the upcoming subsections.

2.4.1. Artificial neural network

ANN is an information-processing algorithm inspired by the human brain's biological neurons (Svozil et al., 1997; Glorot and Bengio, 2010; Zamri et al., 2022; Chen et al., 2023). It mimics and processes the information similar to how nerve cells do in the biological brain. Feed-forward ANN is one of the most successful ANN where the connections of the neurons do not form any loop, and the information propagates in one direction from input to output. The output of neurons is the dot product of inputs (x) and its corresponding weights (w) combined by a summation function (i.e., $w \cdot x$) to which a threshold or bias is added (Eq. (6)).

$$f = \sum w \cdot x + Bias \quad (6)$$

Finally, the output (f) of each neuron will pass through an activation function which decides whether the outcome of any neuron is helpful in predicting the response variable or not. In ANN, we have two primary hyperparameters: weights and biases. A details of the ANN for groundwater applications can be found in Wunsch et al. (2021).

2.4.2. Support vector regression

Support vector regression (SVR) is an extension of the massive kernel range strategies utilised for regression analysis (Vapnik et al., 1996). It preserves all the attributes that describe maximum support vector machine (SVM) algorithms, like entropy, duality, sparsity, and kernel. SVR is a potential tool for predictive data analysis and can be used in several applications like image denoising, voice conversion, and filter design. SVR has recently attracted wireless communication to address various issues like predicting the average localisation error (Singh et al., 2020), antenna selection (Naem and Lee, 2011), decision making (Wu et al., 2012), pilot design (Gao et al., 2014) and high-MIMO beamforming (Wang et al., 2014). It is widely used for both linear and non-linear mapping and prediction using various kernels. The mapping function between input and output is given by

$$f_i = w\zeta(x) + \alpha \quad (7)$$

where x is the input ($x^1, x^2, \dots, x^{N_{train}}$), f_i is the output, $w \in \mathbb{R}$ is the weight vector, α is constant, N_{train} is the number of training data, and $\zeta(\cdot)$ is the non-linear function. The value of w is estimated by using Eq. (8).

$$\begin{aligned} \text{Minimise : } & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N_{train}} (\gamma_i - \gamma_i^*) \\ \text{Subject to : } & \begin{cases} f_i - (w\zeta(x_i) + \alpha) \leq \epsilon + \gamma_i \\ (w\zeta(x_i) + \alpha) - f_i \leq \epsilon + \gamma_i^* \\ \gamma_i, \gamma_i^* \geq 0 \end{cases} \end{aligned} \quad (8)$$

where γ_i, γ_i^* are slack variables, C is the box constraint, and ϵ is the insensitive loss function. Finally, the output can be calculated by using Eq. (9).

$$f_i = \sum_{i=1}^{N_{train}} (\gamma_i + \gamma_i^*) K(x_i, x) + \alpha \quad (9)$$

Algorithm 1 AutoML-GWL for groundwater level prediction

```

1: Inputs: Dataset (Input features) and Column (Response variable) ▷ Input statement
2: Outputs: best_model (Best model) and best_hyperparams (Best hyperparameters) ▷ Output statement
3: function AUTOML-GWL(data : Dataset, target_column : Column)
4:   best_model ← None
5:   best_score ← -∞
6:   train_data, val_data ← split_dataset(data, split_ratio = 0.7)
7:   algorithms ← [ANN, SVR, GPR, RF, Boosting EL, BDT, LR, KR]
8:   for algorithm ∈ algorithms do
9:     best_hyperparams ← bayesian_hyperparameter_tuning(algorithm, train_data, target_column)
10:    score_sum ← 0
11:    for fold ∈ cross_validation_folds(train_data, num_folds) do
12:      model ← algorithm.train(fold.train_data, target_column, hyperparameters = best_hyperparams)
13:      fold_score ← algorithm.evaluate(fold.val_data, target_column)
14:      score_sum += fold_score
15:    end for
16:    avg_score ← score_sum/num_folds
17:    if avg_score > best_score then
18:      best_score ← avg_score
19:      best_model ← model
20:    end if
21:  end for
22:  return best_model
23: end function
24: function BAYESIAN_HYPERPARAMETER_TUNING(algorithm, train_data, target_column)
25:   hyperparameter_space ← algorithm.get_hyperparameter_space()
26:   function OBJECTIVE_FUNCTION(hyperparams)
27:     model ← algorithm.train(train_data, target_column, hyperparameters = hyperparams)
28:     score ← algorithm.evaluate(val_data, target_column)
29:     return score
30:   end function
31:   best_hyperparams ← bayesian_optimisation(objective_function, hyperparameter_space)
32:   return best_hyperparams
33: end function
34: function BAYESIAN_OPTIMISATION(objective_function, hyperparameter_space)
35:   best_hyperparams ← optimise(objective_function, hyperparameter_space)
36:   return best_hyperparams
37: end function

```

where $K(x_i, x)$ is the kernel function. The two main hyperparameters used in SVR are C and ϵ .

2.4.3. Gaussian process regression

GPR is a non-parametric machine learning model based on Bayesian theory (Rasmussen et al., 2006). It is mainly used to resolve complex regression based machine learning problems such as intrusion detection (Singh et al., 2021b), batteries quality prediction (Liu et al., 2019), and wind speed prediction (Cai et al., 2020). The widely used variant of GPR assumes a joint multivariate normal distribution with zero mean and calculates the co-variance by using the squared exponential co-variance function ($K(x, x')$) given by Eq. (10).

$$K(x, x') = \sigma^2 \exp\left[-\frac{\tau}{2}\right] \quad (10)$$

where σ is the model noise and τ is given by Eq. (11).

$$\tau = \frac{|x - x'|^2}{l_s^2} \quad (11)$$

where l_s is the length scale that tunes the vertical and horizontal scale of the function. The two primary hyperparameters in GPR are sigma and length scale.

2.4.4. Ensemble learning

Perrone and Cooper (1995) developed a broad conceptual framework for leveraging ensemble approaches to achieve notably higher regression accuracy. By developing and bringing together a number

of base (or weak) learners with distinct methods, ensemble learning (EL) improves overall performance. It is widely preferred when there is a shortage of training data. It becomes very challenging to select a classifier limited data scenario. By averaging the output of different classifiers, ensemble algorithms reduce the possibility of choosing an inadequate classifier. In this study, we opted for bagging and boosting EL techniques for developing ensemble learning algorithms.

Bagging EL (also known as random forest or bootstrap aggregation method) is proposed by Breiman (1996). It is one of the most popular regression algorithms that uses bootstrapping to randomly create a forest of decision trees based on different training sets (say D training sets). It considers the output from each decision tree in generating the final output through averaging given by Eq. (12).

$$\hat{f}_{average} = \frac{1}{D} \sum_{i=1}^D \hat{f}^D(x) \quad (12)$$

where $\hat{f}^D(x)$ is the individual prediction from D training sets. There are three primary hyperparameters in the case of bagging EL: minimum leaf size of the tree, number of predictors in each split, and number of trees. The first two parameters define the tree morphology, and the third defines the accuracy and efficiency.

The other type of EL is boosting, which is motivated by Kearns and Valiant (1994) and given by Freund et al. (1996). It is slightly different from bagging EL, as in Boosting EL, the bootstrap sampling technique is not used. Instead, the models are created repeatedly and sequentially;

hence the knowledge of model j is required before the model $j + 1$ is generated.

2.4.5. Binary decision tree

BDT is a structure that is based on a sequential decision-making process (Shlien, 1990; Artime Ríos et al., 2019). It performs regression through a recursive binary split on features depending upon whether $x_i \leq \alpha$ or $x_i \geq \alpha$, where $\alpha \in \mathbb{R}$ represent the observed values in BDT given by Eq. (13).

$$DT(x) = \sum_{j=1}^{T_n} j \cdot B_f(x) \quad (13)$$

where $DT(x)$ represent the regression tree, T_n is the terminal nodes of $DT(x)$, and $B_f(x)$ represents the base function given by Eq. (14)

$$B_f(x) = \prod_{i=1}^{T_s} [x_i(j) - sv_{im}] \quad (14)$$

where T_s is the total number of splits, x_i is the enhanced variable, and sv_{im} is the splitting value. A feature is assessed from the root, and one of the two branches is chosen. This process is repeated until the last leaf is reached (i.e., minimum leaf size). Hence, minimum leaf size is an important hyperparameter that needs to be optimised for better performance.

2.4.6. Linear regression

LR model is the most widely used regression method that relates the response variable with input features (Poole and O'Farrell, 1971). A LR for T number of input features and N_{obs} number of observations is given by Eq. (15).

$$\begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_{N_{obs}} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{T1} \\ x_{12} & x_{22} & \dots & x_{T2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1N_{obs}} & x_{2N_{obs}} & \dots & x_{TN_{obs}} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{N_{obs}} \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{N_{obs}} \end{bmatrix}$$

or $f = X\beta + \alpha$ (15)

where the value of β is calculated by using least square method given by Eq. (16).

$$\hat{\beta} = (X^T X)^{-1} X^T f \quad (16)$$

Any prediction say at $x = x_o$ through multivariate linear regression is given by

$$\hat{f} = x_o \cdot \hat{\beta} = x_o (X^T X)^{-1} X^T f \quad (17)$$

2.4.7. Kernel regression

Kernel regression is a non-parametric regression model which solves regression problems by placing kernels (i.e., weighted functions) at all the observation points (Mack and Silverman, 1982; Hart and Wehrly, 1986; Liu et al., 2021). The role of the kernel is to assign a weight to each location depending on the distance between the location and the observation point. Mathematically, a multi-variate kernel regression model relating the response variable (f) and the input (x_i) can be represented by Eq. (18).

$$E(f_i | x_i) = f = \zeta(x_i) + \alpha_i \quad (18)$$

where ζ is the non-linear mapping function, α indicates the unexplained variation in the output around its mean (Eq. (19)) and is independent of x_i .

$$\zeta(x_i) = E[f_i | x_i = x] = \frac{\int f \cdot f(x, f) df}{\int f(x, f) df} = \frac{\int f \cdot f(x, f) df}{\int f(x)} \quad (19)$$

2.5. Bayesian optimisation

Bayesian optimisation directly employs the Bayes theorem to obtain the maxima or minima of the objective function. Unlike random and grid search optimisation techniques, BO drastically improves the search speed by considering previous performances. It makes use of past observation to develop a probabilistic model (also known as a surrogate function) of the objective function. The most often used surrogate function for BO is GP (Eq. (20)).

$$f(x) \sim \zeta(\mu(x), C_v(x, x')) \quad (20)$$

where $f(x)$ is GP distributed function with mean $\mu(x)$ and covariance $C_v(x, x')$. The typically used covariance function is given by Eq. (21)

$$C_v(x, x') = \exp\left(\frac{-1}{2} \|x - x'\|^2\right) \quad (21)$$

The surrogate function is relatively easier to optimise than the original objective function. BO uses the acquisition function to assign the next set of hyperparameters for the original objective function by selecting the parameters that perform best on the surrogate function.

2.6. Working of AutoML-GWL

Commencing with dataset preparation for groundwater level prediction tasks, the AutoML-GWL process involves splitting the data into distinct training (70%) and validation (30%) sets. Within this framework, AutoML-GWL systematically explores an array of regression algorithms, each equipped with its unique set of hyperparameters for fine-tuning. The crux of the approach lies in conducting hyperparameter tuning through BO. This iterative process entails training models with varying hyperparameter combinations and assessing their performance on the validation data. The process identifies optimal hyperparameters by leveraging the achieved validation scores. Subsequently, the algorithm with the highest validation score is singled out. The selected algorithm's model is then trained using its optimal hyperparameters and evaluated on the validation dataset. This cycle iterates across all considered algorithms. Ultimately, the outcome of this intricate process is the identification of the model that exhibits the highest validation performance (i.e., least error) among all assessed algorithms. In summary, the process optimises the hyperparameters of different regression algorithms using BO and selects the algorithm that produces the best predictive performance on the validation data. This approach automates the selection and fine-tuning of algorithms, reducing the need for manual intervention and expertise in algorithm selection and tuning.

3. Results

3.1. Feature importance

We utilise the potential of the regression tree ensemble to evaluate the relevancy of the input features (Fig. 4a). We found that out of the six input features (precipitation, temperature, evaporation, soil type, relative humidity, and GWL lag), GWL lag emerged as the most relevant input feature with the highest relative importance score. It is followed by relative humidity, precipitation, temperature, and evaporation. Soil type has the least importance score, which indicates that it is of less importance in predicting the groundwater level. We also employed the backward elimination approach to assess whether the least relevant feature could be considered redundant. This method involved considering and eliminating the feature with the lowest importance score and then computing the model loss. Surprisingly, we observed that the model loss increased significantly (from 0.755 to 0.845) without the least relevant feature. This result indicated that, despite its relatively low importance score, the feature played a crucial role in learning the dynamics of groundwater level. Hence, it became evident that even the

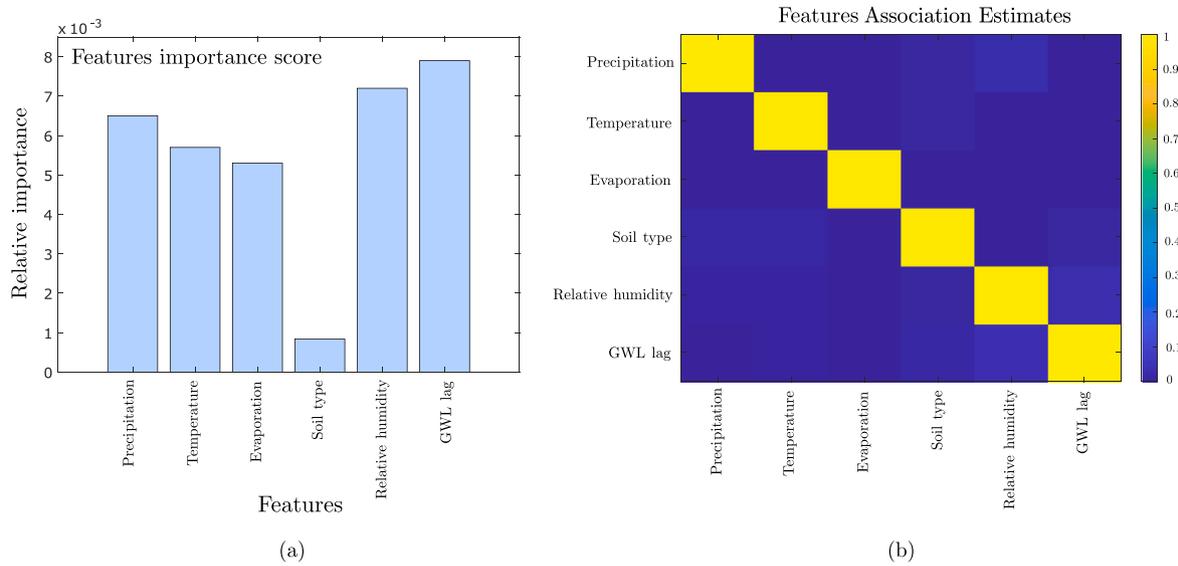


Fig. 4. Feature (a) importance score, and (b) association estimates.

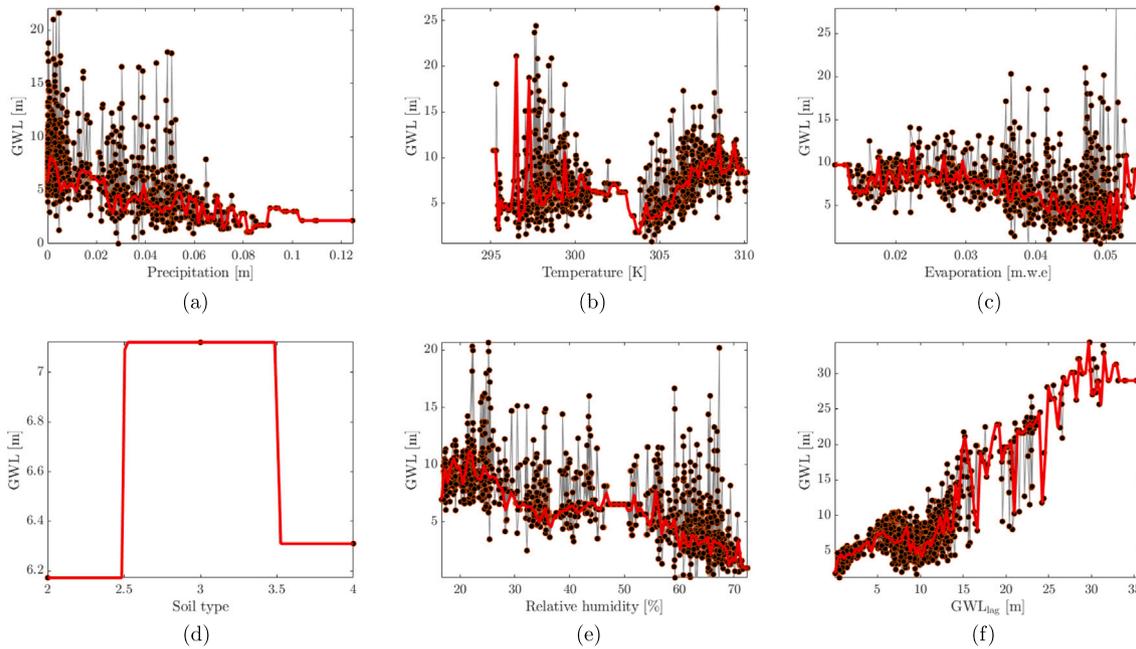


Fig. 5. Feature sensitivity analysis of input features (a)–(f) using partial dependency plot (red line) and individual conditional expectation curve (gray lines).

least important feature proved to be non-redundant and meaningful in the context of our model.

In addition to conducting a feature importance analysis, we also calculated feature association estimates to assess correlations among the input features. The presence of correlated features can lead to model instability, increasing susceptibility to uncertainty. The pixel magnitude in this context reflects the similarities in decision rules for splitting at each observation (refer to Fig. 4b). A high value between any two features indicates a strong correlation. However, our analysis revealed that there were no significant correlations among the input features.

3.2. Feature sensitivity

We conducted a rigorous feature sensitivity analysis of the input feature to delineate the individual impact (*i.e.*, positive, negative, or

undulating) of each feature on the response variable. We found an obvious negative impact of precipitation on GWL (Fig. 5a). The temperature shows an undulating behaviour which can be separately analysed in two clusters (Fig. 5b). The first cluster ranges from 295 K to 302 K, and the second cluster ranges from 303 K to 310 K. The first cluster corresponds to the post-monsoon period, whereas the second cluster corresponds to a mixture of pre-monsoon and monsoon. Both these clusters show a positive impact on GWL. We observed a mixed response in the case of evaporation, where the GWL corresponding to monsoon (>0.045) is less (Fig. 5c). In contrast, the GWL is high during the post-monsoon (usually between 0.05–0.045). The presence of small undulation in the PDP plot mark the interference of the pre-monsoon evaporation data with the monsoon and post-monsoon data. We do not observe much variation due to soil type; however, medium fine texture soil (*i.e.*, Soil 3) have usually high GWL (Fig. 5d). Relative humidity has an overall negative response which can be segregated into

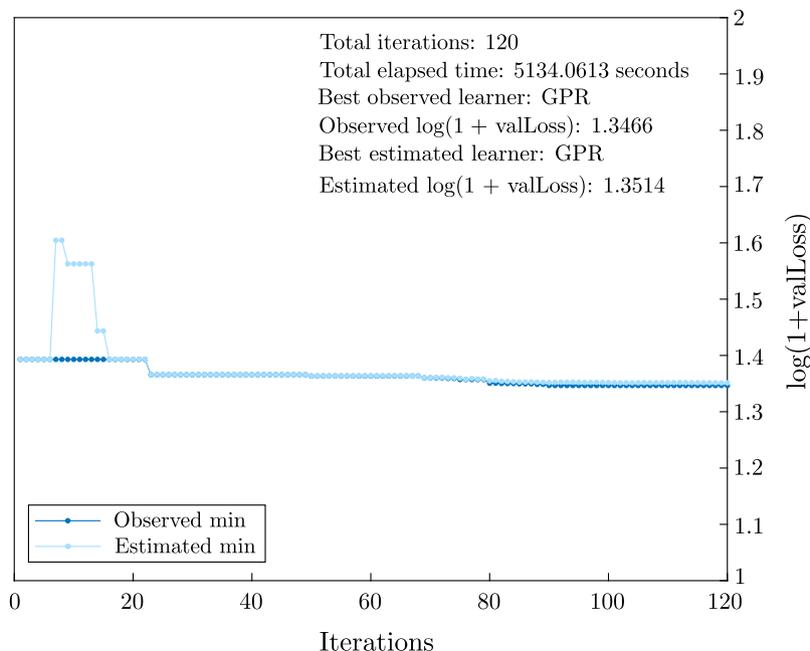


Fig. 6. Best cost curve illustrating the training loss during the selection of the optimal machine learning process by Bayesian optimisation.

three clusters. The first cluster (<32%) corresponds to pre-monsoon, the second cluster (between 32%–55%) corresponds to post-monsoon, and the third cluster (>55%) corresponds to the third cluster. The first and the third cluster show a negative impact, whereas the second cluster shows a steady impact (Fig. 5e). We observe an obvious positive impact of GWL_{lag} on GWL (Fig. 5f).

3.3. Performance of the AutoML-GW

We divided all the input features, along with the corresponding response variable, into 70% for training and 30% for testing the AutoML-GWL model. We fed training data into the AutoML-GWL model to select the best-performing algorithm for groundwater level prediction. AutoML-GWL iteratively optimised the participating algorithms' hyperparameters by using Bayesian optimisation (Fig. 6). We set the maximum iteration to 120. At each iteration, it computes the fitness value from the fitness function (Eq. (22)) by leveraging any one of the participating algorithms.

$$Fitness = \log(1 + Loss_{validation}) \quad (22)$$

where $Loss_{validation}$ is the mean square error involved in the cross-validation. After the completion of each iteration, the model returns the optimised hyperparameters along with the fitness value. After the completion of a maximum iteration, the AutoML-GWL model returns the model corresponding to the lowest fitness value. We found that GPR exhibits the lowest fitness value at iteration number 90. Hence, the AutoML-GWL returns GPR as the best-performing algorithm model along with the optimised hyperparameters (i.e., $\sigma = 1.478$). Before returning, the AutoML-GWL model retrained the GPR model on the complete training datasets.

After the selection of the best-performing algorithm by the AutoML-GWL framework, we evaluated the training accuracy of the GPR by computing the model performance on the training datasets. We found that the model exceptionally well on the training datasets ($N = 5817$) with $R = 0.91$, $RMSE = 1.51$, and $bias = 0$. The training accuracy confirms that the model perfectly captures the complexity of the datasets. However, for a fair assessment, we assessed the model performance on the unseen datasets (i.e., testing data). We found the trained model retains its training level accuracy on the testing data with negligible generalisation error. We plotted a linear fit between the predicted and

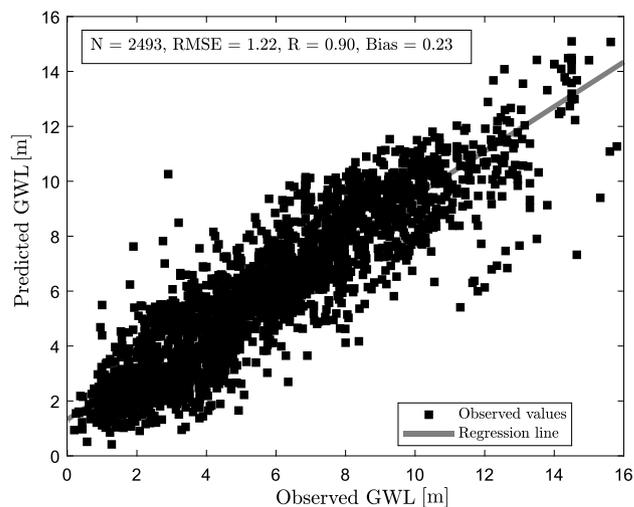


Fig. 7. Regression curve between the unseen observed and AutoML-GWL predicted groundwater level.

observed groundwater levels (Fig. 7). We found that the maximum observations are centric along the regression line with $R = 0.90$, $RMSE = 1.22$, and $bias = 0.23$.

3.4. Error histogram analysis

We conducted an error histogram analysis to thoroughly inspect the distribution of errors associated with our proposed model. To begin, we calculated the testing errors by subtracting the observed values from the predicted ones. Subsequently, we plotted the histogram of the errors using twenty bins (Fig. 8). Our analysis revealed that the total testing error symmetrically ranged from -6.95 (first bin on the left) to 6.99 (last bin on the right). The negative errors situated to the left of the zero error line indicated underestimation errors, while the positive errors on the right side indicated overestimation errors. Additionally, we fitted a Gaussian curve to the histogram and found that the errors followed

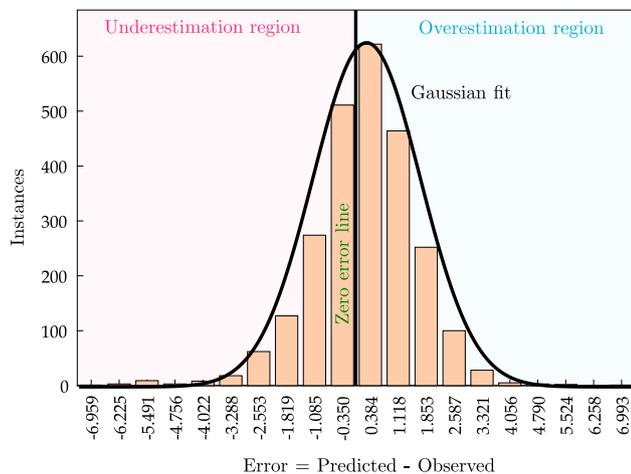


Fig. 8. Error pattern evaluation using error histogram analysis considering twenty bins. The green vertical line indicates the zero error line, which bifurcates the error space into two regions: underestimation (left to the green line) and overestimation (right to the green line).

a normal distribution, with the peak centred near the zero error line (indicated in green) in the overestimation region. This observation suggested that our model exhibited low bias, as the majority of the errors were clustered closely around zero. The presence of minimal bias was further supported by the low positive bias value of 0.23.

The error histogram analysis provided valuable insights into the performance of our model and the distribution of errors across the testing data. The symmetric distribution of errors and the close alignment of the peak with the zero error line indicated the model's ability to achieve accurate predictions with minimal systematic deviations. These findings contribute to the credibility of our proposed model, demonstrating its suitability for accurate groundwater level prediction and providing confidence in its practical applicability. The low bias and normally distributed errors are strong indicators of the model's reliability, enhancing its potential for various real-world applications in spatial analysis and groundwater level forecasting.

3.5. Residual analysis

We performed residual analysis to assess the appropriateness of the fitted model. The residuals were computed by subtracting the in-situ observations from the corresponding fitted values (Fig. 9). We observed that the residuals were uniformly scattered above and below the zero residual line, confirming their stochastic nature. Additionally, we did not observe any discernible trend or pattern in the residuals, indicating that all six input features comprehensively explained the groundwater level variations. Ideally, the mean of all the residuals should be zero, and upon analysis, we found that the mean was approximately zero ($\mu_{residual} = 2.85E-17$). This result supports the accuracy of our model and suggests that the residuals are well-distributed around zero, reflecting the model's capability to capture the underlying patterns in the data effectively.

Furthermore, we evaluated the independent nature of the residuals by computing the serial correlation between the residual and its lagged version. The correlation coefficient between them was very low ($R < 0.03$), indicating that the machine learning model extracted the maximum information from the datasets. This lack of significant correlation signifies that the model successfully accounted for temporal dependencies and was not influenced by any hidden patterns or biases. By conducting a thorough residual analysis, we verified the validity and reliability of our fitted model. The stochastic nature of the residuals, absence of trends, and minimal serial correlation demonstrate the model's ability to accurately represent the groundwater level variations based

on the six input features. The results of the residual analysis strengthen the credibility of our research findings and highlight the model's capability to effectively capture the complexities of groundwater level prediction. This analysis contributes valuable insights, providing a solid foundation for the practical application and broader understanding of our proposed machine learning approach.

4. Discussion

4.1. Comparison with benchmark algorithms

We conducted a comprehensive performance evaluation of AutoML-GWL, comparing it with sixteen widely-used benchmark algorithms for monitoring and predicting groundwater levels. The benchmark algorithms encompassed both conventional baseline machine learning models such as RF, Boosting EL, BDT, GAM, GRNN, LR, ANN, SVR, RBNN, KR, and LSTM, as well as novel hybrid machine learning models based on fuzzy and meta-heuristic algorithms. For the hybrid models, we integrated Fuzzy Inference Systems (FIS) with various metaheuristic algorithms, including Teaching-Learning Based Optimisation (TLBO), Ant Colony Optimisation (ACO), Harmony Search (HS), Differential Evolution (DE), and Weevil Damage Optimisation Algorithm (WDOA). This combination allowed us to accurately map groundwater levels by leveraging the strengths of both fuzzy logic and metaheuristics. All the benchmark algorithms were trained and evaluated on the same datasets, as presented in Tables 2 and 3. To assess the performance of these models, we employed three performance metrics: correlation coefficient (R), root mean square error (RMSE), and bias.

Our results revealed that AutoML-GWL outperformed all the benchmark algorithms, achieving the highest correlation coefficient ($R = 0.90$) and the lowest error ($RMSE = 1.22$). These findings highlight the superior predictive capabilities of AutoML-GWL in capturing the underlying relationships and patterns within the groundwater level data. Notably, SVR demonstrated the lowest bias of -0.01 , indicating its minimal tendency to systematically over or underestimate the groundwater level. On the other hand, KR exhibited the lowest correlation coefficient ($R = 0.64$) and the highest bias (0.77), indicating its limited performance in accurately representing groundwater level variations. Moreover, Fuzzy-DE showed the highest error ($RMSE = 2.67$), indicating higher prediction inaccuracies than the other models. Our study demonstrates that AutoML-GWL is a highly effective approach for groundwater level prediction, surpassing traditional benchmark algorithms and novel hybrid models. Its exceptional performance makes it a promising tool for groundwater monitoring and management applications.

For a more robust analysis, we performed a statistical test to check whether the performance of all the models was statistically different or the same. To do so, we first calculated the error of all the benchmark algorithms and the AutoML-GWL. We applied Kolmogorov-Smirnov (KS) test to perform the normality test of the error data. We found that the errors of all seventeen models are normally distributed. We then perform one-way ANalysis Of VAriance (ANOVA) to test the null hypothesis that all the models have the same mean or not, which is the alternate hypothesis (Fig. 10). We reject the null hypothesis as eleven models (Boosting EL, GRNN, LR, ANN, SVR, KR, LSTM, Fuzzy-TLBO, Fuzzy-ACO, Fuzzy-DE, and Fuzzy-WDOA) are significantly different from AutoML-GWL in terms of mean (p -value < 0.05). The overlapping of the comparison interval of RF, BDT, GAM, RBNN, and Fuzzy-HS with the comparison interval of the AutoML-GWL indicates that the means of these algorithms are not significantly different from each other. In addition to performing ANOVA, we also applied the Friedman test to calculate the mean rank of all the benchmark algorithms (Table 4). The Friedman test is a non-parametric statistical test that evaluates the performance of algorithms (Jamaludin et al., 2022; Kasihmuddin et al., 2022). Lower ranks in the Friedman test indicate better overall

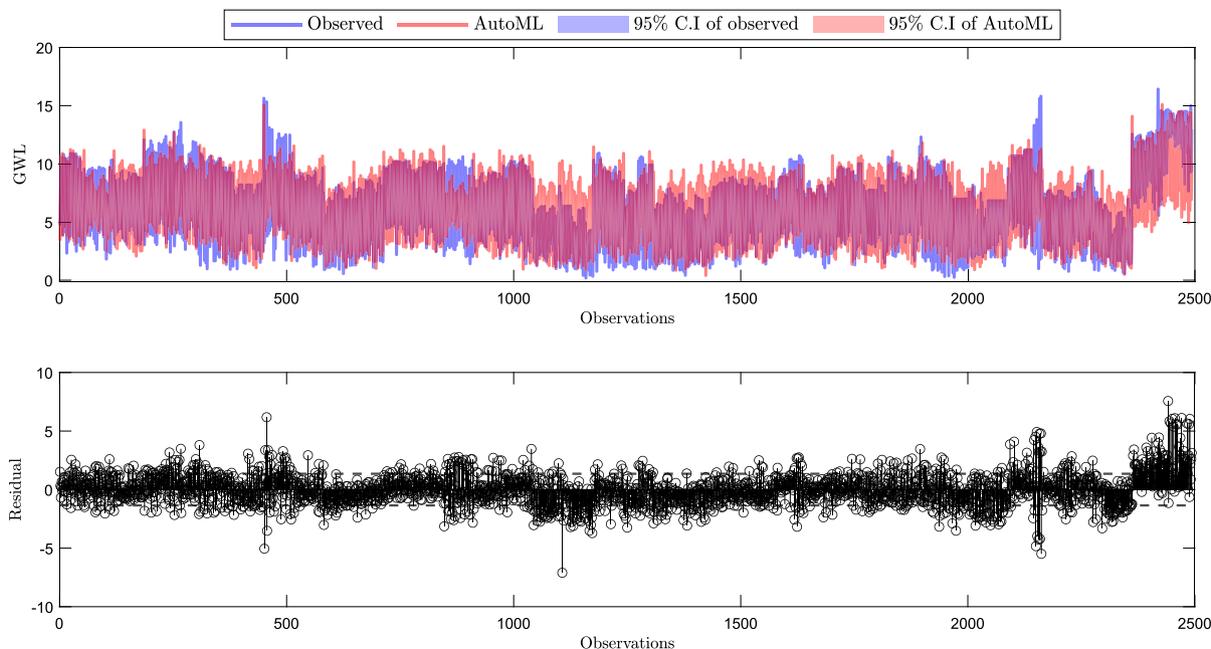


Fig. 9. The top panel shows the line plot between the observed and predicted groundwater levels. The bottom panel shows the residual between the observed and fitted groundwater levels. The dashed line above and below the zero residual line indicates \pm testing RMSE.

Table 2

Comparison of the AutoML-GWL results with benchmark algorithms. The best and worst values in each row are marked in blue and red, respectively.

Performance metrics	Random forest	Boosting EL	BDT	GAM	GRNN	Linear regression	Artificial neural network	SVR	RBNN	Kernel regression	LSTM	AutoML-GWL (This study)
R	0.88	0.83	0.82	0.89	0.89	0.88	0.85	0.88	0.87	0.64	0.86	0.90
RMSE	1.26	1.72	1.81	1.31	1.32	1.32	1.96	1.31	1.34	1.51	1.31	1.22
Bias	0.31	0.38	0.30	0.13	0.02	-0.05	0.02	-0.01	0.09	0.77	0.41	0.23

Table 3

Comparison with novel metaheuristic-based benchmark algorithms.

Performance metrics	Fuzzy-TLBO	Fuzzy-ACO	Fuzzy-HS	Fuzzy-DE	Fuzzy-WDOA
R	0.87	0.88	0.83	0.77	0.86
RMSE	1.74	1.80	2.26	2.67	1.90
Bias	0.23	0.04	-0.16	0.59	-0.03

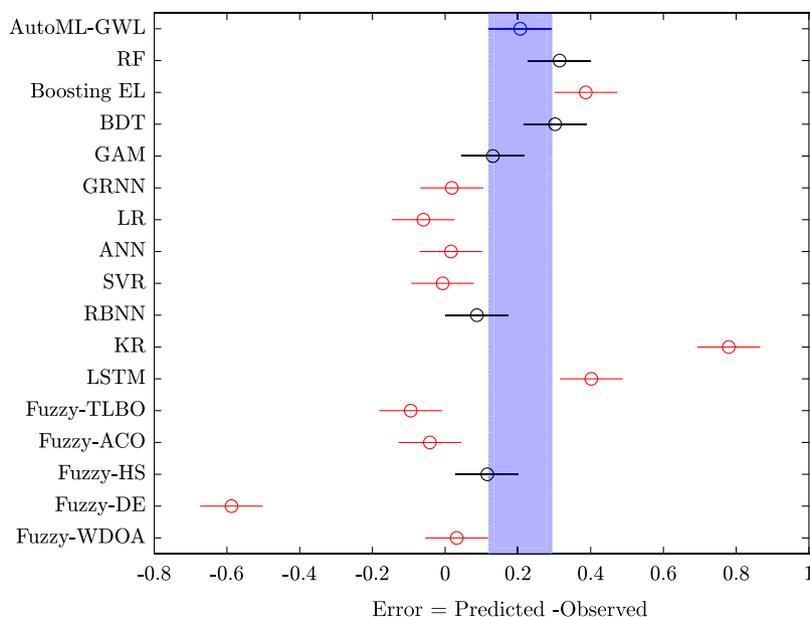


Fig. 10. Statistical analysis of all the benchmark algorithms. The vertical blue strip indicates the comparison interval of AutoML-GWL.

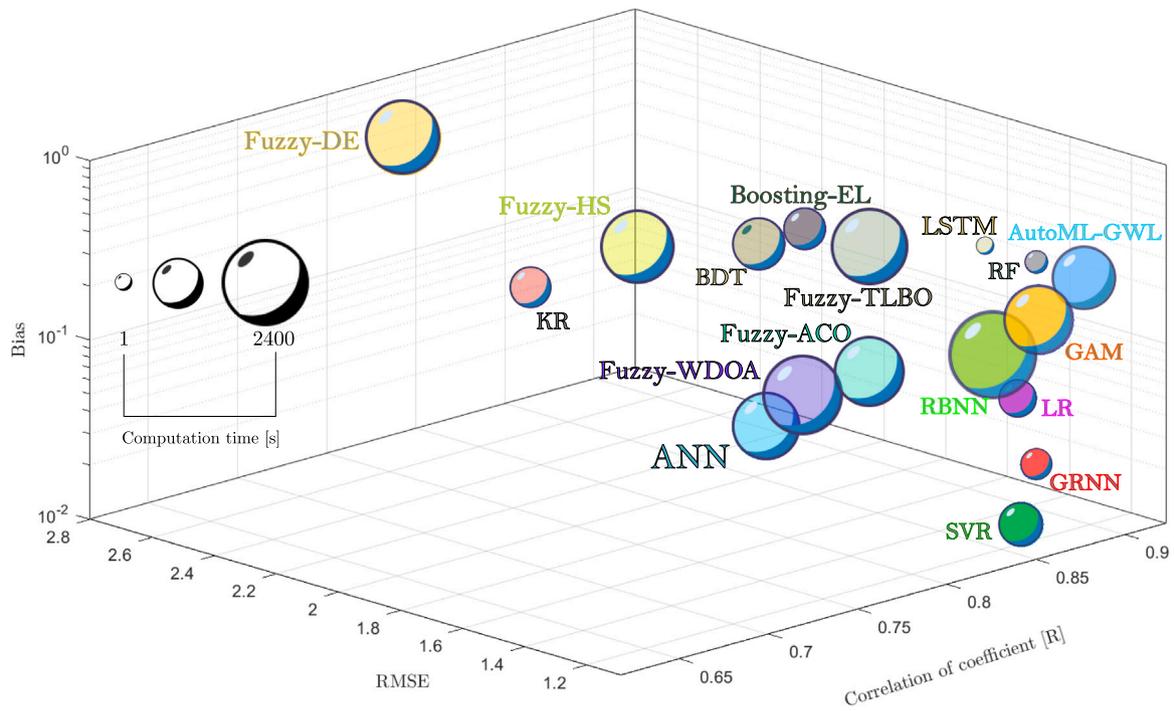


Fig. 11. Bubble plot illustrating the computational time along with the other performance metrics (i.e., R, RMSE, and Bias). The size of the circle represents the computation time of each algorithm. The z-axis is in the log scale.

Table 4
Friedman rank test on all the benchmark algorithms. The best value is marked in blue colour.

Algorithms	RF	Boosting EL	BDT	GAM	GRNN	LR	ANN	SVR	RBNN	KR	LSTM	AutoML-GWL	Fuzzy-TLBO	Fuzzy-ACO	Fuzzy-HS	Fuzzy-DE	Fuzzy-WDOA
Mean rank	3.8	11.8	14.0	3.3	4.5	6.0	13.5	4.8	8.3	13.0	7.3	1.0	9.8	8.8	14.8	16.5	12.3



Fig. 12. Uncertainty analysis of the AutoML-GWL model in presence of $\pm 5\%$ and $\pm 10\%$ uncertainties in each input features.

performance among the tested algorithms. We found that the AutoML-GWL achieved the first rank among all the benchmark algorithms, indicating its superior performance.

Apart from analysing accuracy-based performance metrics (i.e., R, RMSE, and bias), it is a common practice in spatial machine learning to also consider the computation time of the algorithms. Computation time usually depends on the algorithm’s efficiency and the size of the input datasets (denoted by n). We recorded the computation time of all

the benchmark algorithms on the same datasets. Notably, we found that the computation time of the AutoML-GWL algorithm exhibits a cubic relationship with the input size (i.e., $\mathcal{O}(n^3)$). In Fig. 11, we present a plot that illustrates the computation time of all the algorithms, along with their corresponding performance metrics. It is worth noting that LSTM and RF demonstrate relatively lower computation times, while RBNN stands out with the highest computation time among the considered

Table 5
Spatial distribution analysis considering thirty different sets of training and testing datasets.

Scenarios	rng(seed)	Training			Testing		
		R	RMSE	Bias	R	RMSE	Bias
Scenario 1	rng(0)	0.910	1.509	1.44E-04	0.899	1.224	0.234
Scenario 2	rng(1)	0.910	1.510	-2.59E-05	0.899	1.222	0.230
Scenario 3	rng(2)	0.917	1.457	-4.02E-05	0.897	1.232	0.201
Scenario 4	rng(3)	0.908	1.522	1.34E-05	0.898	1.230	0.228
Scenario 5	rng(4)	0.916	1.468	2.84E-05	0.898	1.227	0.202
Scenario 6	rng(5)	0.914	1.479	7.58E-05	0.899	1.219	0.211
Scenario 7	rng(6)	0.909	1.519	-6.22E-05	0.898	1.227	0.228
Scenario 8	rng(7)	0.917	1.459	1.71E-05	0.897	1.234	0.199
Scenario 9	rng(8)	0.909	1.516	-5.79E-05	0.898	1.225	0.228
Scenario 10	rng(9)	0.911	1.501	-4.19E-06	0.900	1.218	0.227
Scenario 11	rng(10)	0.918	1.455	5.05E-05	0.896	1.238	0.197
Scenario 12	rng(11)	0.910	1.512	5.63E-05	0.900	1.223	0.232
Scenario 13	rng(12)	0.910	1.507	1.47E-05	0.899	1.220	0.230
Scenario 14	rng(13)	0.909	1.514	5.18E-05	0.899	1.225	0.232
Scenario 15	rng(14)	0.908	1.524	2.99E-05	0.898	1.231	0.228
Scenario 16	rng(15)	0.908	1.522	-3.91E-05	0.898	1.229	0.226
Scenario 17	rng(16)	0.909	1.513	1.55E-05	0.899	1.223	0.231
Scenario 18	rng(17)	0.908	1.521	5.86E-05	0.898	1.228	0.228
Scenario 19	rng(18)	0.908	1.525	1.04E-05	0.898	1.230	0.224
Scenario 20	rng(19)	0.908	1.520	2.65E-05	0.898	1.228	0.228
Scenario 21	rng(20)	0.909	1.518	9.52E-05	0.898	1.227	0.231
Scenario 22	rng(21)	0.911	1.504	-1.55E-05	0.900	1.219	0.229
Scenario 23	rng(22)	0.912	1.499	-6.04E-05	0.900	1.220	0.223
Scenario 24	rng(23)	0.908	1.520	4.22E-05	0.898	1.227	0.228
Scenario 25	rng(24)	0.909	1.519	2.64E-05	0.898	1.229	0.230
Scenario 26	rng(25)	0.909	1.513	1.81E-05	0.899	1.225	0.231
Scenario 27	rng(26)	0.908	1.522	-1.05E-05	0.898	1.229	0.227
Scenario 28	rng(27)	0.909	1.518	-3.11E-05	0.898	1.227	0.228
Scenario 29	rng(28)	0.908	1.522	6.32E-05	0.898	1.230	0.229
Scenario 30	rng(29)	0.909	1.517	7.45E-05	0.898	1.227	0.231
Mean \pm standard deviation		0.910 \pm 0.003	1.507 \pm 0.021	1.89E-05 \pm 4.78E-05	0.898 \pm 0.001	1.226 \pm 0.005	0.224 \pm 0.010

algorithms. AutoML-GWL, despite being accurate, is computationally expensive.

4.2. Uncertainty analysis of the AutoML-GWL

We conducted a comprehensive uncertainty analysis to assess the stability of the proposed AutoML-GWL model. This analysis involved quantifying the variability in the model's output based on the variations in the input features. To systematically examine the impact of uncertainties, we deliberately introduced small fluctuations ($\pm 5\%$ and $\pm 10\%$) in each input feature while keeping all other inputs constant. In Fig. 12, we presented the mean percentage change in the response variable resulting from these small uncertainties. Our findings reveal that the AutoML-GWL model demonstrates remarkable stability, with changes in the output ranging from -4.66% to $+4.64\%$ in response to the small input uncertainties. Notably, we identified relative humidity, the second most relevant feature, emerges as the most prone feature which is highly susceptible to get affected by small uncertainties as compared to other features. The output variability approximately doubled when the uncertainty in relative humidity varied from $\pm 5\%$ to $\pm 10\%$.

In contrast, temperature, soil type, and evaporation were identified as the most stable features, with minimal variability in the output. The maximum percentage change in the output corresponding to temperature, soil type, and evaporation was found to be -0.12% , -0.10% , and 0.29% , respectively. Additionally, GWL lag showed strong resistance to small uncertainties, with changes in the response variable remaining below 1% . Regarding the uncertainty in precipitation, we observed variations in the output ranging from -1.08% to 1.05% . Overall, our analysis demonstrates that the proposed AutoML-GWL model exhibits considerable stability in response to input uncertainties. These findings validate the reliability and robustness of our model, providing greater confidence in its performance across various real-world scenarios.

The inclusion of this uncertainty analysis provides valuable insights into the behaviour of our model, significantly bolstering the practical

applicability and robustness of our research. This in-depth examination of uncertainty not only enhances our understanding of the model's performance but also instills greater confidence in its real-world utility. By assessing the model's stability under varying input conditions, we establish its resilience and suitability for handling data from different geographical regions and varying environmental factors. The observed robustness in the face of small input perturbations indicates the model's ability to generalise effectively, making it an indispensable tool for groundwater level prediction tasks in various contexts. By addressing and validating the model's response to uncertainties, we pave the way for wider adoption and utilisation of the proposed AutoML-GWL model in practical applications, benefiting researchers, practitioners, and stakeholders alike.

4.3. Spatial distribution analysis

In spatial machine learning, assessing the dependency of training and testing datasets on local factors, such as topological, geometric, or geographic properties, is of utmost importance. Ideally, datasets should exhibit spatial independence to ensure reliable model performance. To evaluate this, we employed the Mersenne Twister (MT) random number generator to create thirty distinct training and testing datasets from the original data. Each scenario utilised a different seed, resulting in thirty independent datasets for both training and testing.

The AutoML-GWL model was retrained on these thirty training sets, and the training and testing accuracies were recorded in Table 5. Remarkably, we observed minimal changes in the performance metrics, indicating the proposed AutoML-GWL model's spatial stability. This finding is further supported by the mean \pm standard deviation values of the training and testing accuracy, which exhibited consistency across the diverse scenarios. The model's spatial stability demonstrated through the consistent performance across various spatially independent datasets instills confidence in its ability to generalise well to different geographic regions or topological variations. This characteristic

enhances the reliability of our spatial machine learning approach and suggests that the AutoML-GWL model is capable of handling spatially diverse data with resilience.

By thoroughly evaluating and reporting on the model's spatial stability, we reinforce the credibility and practical applicability of our research findings. The assessment of spatial independence serves as a crucial validation step, ensuring that the model's performance is not unduly influenced by specific local factors. This aspect contributes significantly to the advancement of spatial machine learning, as it enhances the model's robustness and generalisability across diverse geographical, topological, or geometric conditions. By addressing the spatial dependency of the training and testing datasets, we provide valuable insights that increase the reliability and real-world relevance of our research. Moreover, this aspect can have broader implications, facilitating the application of our approach to a wider range of spatial analysis tasks and fostering the development of more reliable and accurate spatial machine learning methodologies.

4.4. Impact analysis of the AutoML-GWL model

Groundwater management and prediction play a vital role in ensuring sustainable water resource management and addressing various environmental challenges. In this subsection, we conduct an impact analysis of the proposed AutoML-GWL model. Conducting an impact analysis is essential as it provides a comprehensive understanding of the real-world implications and potential benefits of the proposed model, ensuring its relevance and practical applicability in addressing current challenges and informing decision-making processes (Muhammad Sidik et al., 2022).

- **Improved Accuracy and Precision:** The AutoML-GWL model's superior predictive capabilities, as demonstrated in the benchmarking results, have significant implications for groundwater level monitoring and prediction accuracy. By providing more precise and accurate predictions, the model can empower water resource managers and decision-makers with reliable information to make informed choices in water allocation and usage.
- **Resource Efficiency:** In addition to its accuracy, the AutoML-GWL model offers enhanced resource efficiency. Its automated machine learning process optimises hyperparameter tuning and model selection, reducing the need for manual intervention and computational resources. This efficiency can result in cost savings and accelerated analysis, making groundwater level prediction more accessible and cost-effective.
- **Robustness and Generalisation:** The robustness of the AutoML-GWL model is a critical aspect of its impact. Through extensive testing and evaluation of diverse datasets, we have demonstrated its ability to perform consistently across various regions and timeframes. This robustness assures users that the model's predictions remain reliable and stable across different groundwater monitoring scenarios.
- **Decision Support Systems:** Accurate and timely groundwater level predictions are essential for effective decision support systems. The AutoML-GWL model's high-performance capabilities make it a valuable tool for developing decision support systems in water management and planning. It can aid in designing proactive water allocation strategies, managing water usage during droughts, and planning for potential environmental impacts.
- **Societal and Environmental Implications:** The accurate prediction of groundwater levels has broader societal and environmental implications. It can facilitate sustainable water resource management, ensuring the availability of safe drinking water and protecting ecosystems that rely on groundwater. By fostering responsible groundwater management practices, the AutoML-GWL model can contribute to mitigating water scarcity and promoting environmental conservation efforts.

- **Enhanced Research and Innovation:** The development and application of the AutoML-GWL model open avenues for further research and innovation in groundwater level prediction. By showcasing the potential of automated machine learning techniques in this domain, our research inspires the exploration of new methodologies and encourages the integration of advanced technologies in groundwater monitoring and management.

In conclusion, the impact analysis of the AutoML-GWL model demonstrates its significance in addressing real-world challenges related to groundwater monitoring and prediction. Its improved accuracy, resource efficiency, robustness, and potential for decision support systems highlight its potential to revolutionise the field of groundwater management. Moreover, the societal and environmental implications underscore its role in sustainable water resource management and ecosystem preservation. The AutoML-GWL model serves as a valuable addition to the toolbox of water resource managers, environmental researchers, and policymakers, providing a reliable and efficient solution for precise groundwater level prediction and informed decision-making.

5. Conclusion

Accurate information on GWL fluctuation directly provides a measure of groundwater resource management. It is very helpful in understanding the dynamics of the aquifer. The purpose of this study is to provide a framework that automatically selects the best performing machine learning model for GWL prediction. We used precipitation, temperature, evaporation, soil type, relative humidity, and lag of GWL as potential input features to train and validate the automated machine learning model (AutoML-GWL) using huge GWL data from 665 wells from duration 1997 to 2018. We found that out of these input features, GWL lag emerged as the most relevant input feature followed by relative humidity and precipitation. We compared the results of the AutoML-GWL with sixteen different benchmark algorithms. The results demonstrate the high prediction accuracy of the proposed model. Further, the uncertainty and spatial analysis of the AutoML-GWL confirm its reliability.

Although we are getting good accuracy through AutoML-GWL, but much remains to be done in this regard. The dependency on real datasets is the major limitation of any machine learning-based framework because real data have many constraints, such as gaps, duplicates, and incoherencies. This problem can be solved by considering synthetic datasets that explain groundwater dynamics. In addition, the concept of synthetic features can be coupled with AutoML-GWL to enhance the accuracy and reliability further. Furthermore, the performance of AutoML-GWL can also be tested using nature-inspired algorithms, which will incorporate the nature-based solution theme.

The outcome of this model can assist managers and planners in accurately planning the region's water resources. It will help the policy-makers to efficiently manage and regulate watersheds and basins. We conclude that automating the hyperparameter tuning and selection of the best algorithm can be used to accurately map the GWL fluctuations which enable groundwater management and ultimately helps the decision-makers to distribute water fairly among stakeholders.

CRediT authorship contribution statement

Abhilash Singh: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Sharad Patel:** Writing – original draft, Writing – review & editing, Visualization. **Vipul Bhadani:** Investigation, Data curation, Visualisation. **Vaibhav Kumar:** Investigation, Writing – review & editing, Visualization, Supervision. **Kumar Gaurav:** Investigation, Writing – original draft, Writing – review & editing, Visualization, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

KG would like to acknowledge IISER Bhopal for providing the institutional support and necessary equipment needed to conduct this research. AS is grateful to the Department of Science and Technology (DST), Government of India, for funding his Ph.D. studies at IISER Bhopal under DST INSPIRE fellowship (Grant No. DST/INSPIRE Fellowship/[IF180001]). SP is supported through the Institute's postdoctoral fellowship at IISER Bhopal. The authors would like to express their gratitude to the editor and the five anonymous reviewers for their valuable comments and suggestions.

Funding

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Code availability

The code is available for download at <https://abhilashsingh.net/codes.html>.

References

- Ahamed, A., Knight, R., Alam, S., Pauloo, R., Melton, F., 2022. Assessing the utility of remote sensing data to accurately estimate changes in groundwater storage. *Sci. Total Environ.* 807, 150635.
- Ahmadi, S.H., Sedghamiz, A., 2007. Geostatistical analysis of spatial and temporal variations of groundwater level. *Environ. Monit. Assess.* 129, 277–294.
- Ahmadi, S.H., Sedghamiz, A., 2008. Application and evaluation of kriging and cokriging methods on groundwater depth mapping. *Environ. Monit. Assess.* 138, 357–368.
- AlOmar, M.K., Hameed, M.M., AlSaadi, M.A., 2020. Multi hours ahead prediction of surface ozone gas concentration: robust artificial intelligence approach. *Atmos. Pollut. Res.* 11 (9), 1572–1587.
- Artme Ríos, E.M., Sánchez Lasheras, F., Suárez Sánchez, A., Iglesias-Rodríguez, F.J., Seguí Crespo, M.d.M., 2019. Prediction of computer vision syndrome in health personnel by means of genetic algorithms and binary regression trees. *Sensors* 19 (12), 2800.
- Azizi, H., Ebrahimi, H., Mohammad Vali Samani, H., Khaki, V., 2021. Evaluating the effects of climate change on groundwater level in the varamin plain. *Water Supply* 21 (3), 1372–1384. <http://dx.doi.org/10.2166/ws.2021.007>, arXiv:<https://iwaponline.com/ws/article-pdf/21/3/1372/886304/ws021031372.pdf>.
- Banadkooki, F.B., Ehteram, M., Ahmed, A.N., Teo, F.Y., Fai, C.M., Afan, H.A., Sapitang, N., El-Shafie, A., 2020. Enhancement of groundwater-level prediction using an integrated machine learning model optimized by whale algorithm. *Nat. Resour. Res.* 29, 3233–3252.
- Bansod, S.D., Yin, Z.-Y., Lin, Z., Zhang, X., 2003. Thermal field over tibetan plateau and Indian summer monsoon rainfall. *Int. J. Climatol.: J. R. Meteorol. Soc.* 23 (13), 1589–1605.
- Bhagat, S.K., Tiyasha, T., Awadh, S.M., Tung, T.M., Jawad, A.H., Yaseen, Z.M., 2021. Prediction of sediment heavy metal at the Australian bays using newly developed hybrid artificial intelligence models. *Environ. Pollut.* 268, 115663.
- Bisoyi, N., Gupta, H., Padhy, N.P., Chakrapani, G.J., 2019. Prediction of daily sediment discharge using a back propagation neural network training algorithm: A case study of the narmada river, India. *Int. J. Sediment Res.* 34 (2), 125–135.
- Breiman, L., 1996. Bagging predictors. *Mach. Learn.* 24, 123–140.
- Buckland, C., Bailey, R., Thomas, D., 2019. Using artificial neural networks to predict future dryland responses to human and climate disturbances. *Sci. Rep.* 9 (1), 3855.
- Cai, H., Jia, X., Feng, J., Li, W., Hsu, Y.-M., Lee, J., 2020. Gaussian process regression for numerical wind speed prediction enhancement. *Renew. Energy* 146, 2112–2123.
- Chambel, A., 2015. The role of groundwater in the management of water resources in the world. *Proc. Int. Assoc. Hydrol. Sci.* 366, 107–108. <http://dx.doi.org/10.5194/piahs-366-107-2015>.
- Chen, J., Kasihmuddin, M.S.M., Gao, Y., Guo, Y., Mansor, M.A., Romli, N.A., Chen, W., Zheng, C., 2023. PRO2SAT: Systematic probabilistic satisfiability logic in discrete hopfield neural network. *Adv. Eng. Softw.* 175, 103355.
- Chidepudi, S.K.R., Massei, N., Jardani, A., Henriot, A., Allier, D., Baulon, L., 2023. A wavelet-assisted deep learning approach for simulating groundwater levels affected by low-frequency variability. *Sci. Total Environ.* 865, 161035.
- Coelho, V.H.R., Montenegro, S., Almeida, C.N., Silva, B.B., Oliveira, L.M., Gusmao, A.C.V., Freitas, E.S., Montenegro, A.A., 2017. Alluvial groundwater recharge estimation in semi-arid environment using remotely sensed data. *J. Hydrol.* 548, 1–15.
- Deo, R.C., Şahin, M., 2015. Application of the artificial neural network model for prediction of monthly standardized precipitation and evapotranspiration index using hydrometeorological parameters and climate indices in eastern Australia. *Atmos. Res.* 161, 65–81.
- Dhawan, S., George, M.P., Jayachandran, K., Khare, M., 2023. Tropospheric ozone variability in Delhi during pre & post monsoon periods: Decoding influence of seasonal variation, diurnal variation, short-range and long-range transport. *Urban Clim.* 47, 101374.
- Evans, P., 2006. Scaling and assessment of data quality. *Acta Crystallogr. D* 62 (1), 72–82.
- Freund, Y., Schapire, R.E., et al., 1996. Experiments with a new boosting algorithm. In: *Icml*, Vol. 96. Citeseer, pp. 148–156.
- Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Ann. Statist.* 1189–1232.
- Gao, Z., Dai, L., Wang, Z., 2014. Structured compressive sensing based superimposed pilot design in downlink large-scale MIMO systems. *Electron. Lett.* 50 (12), 896–898.
- Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. In: *JMLR Workshop and Conference Proceedings*, pp. 249–256.
- Goldstein, A., Kapelner, A., Bleich, J., Pitkin, E., 2015. Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *J. Comput. Graph. Statist.* 24 (1), 44–65.
- Guyon, I., Sun-Hosoya, L., Boullé, M., Escalante, H.J., Escalera, S., Liu, Z., Jajetic, D., Ray, B., Saeed, M., Sebag, M., et al., 2019. Analysis of the autotml challenge series. *Autom. Mach. Learn.* 177.
- Hameed, M., Sharqi, S.S., Yaseen, Z.M., Afan, H.A., Hussain, A., Elshafie, A., 2017. Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in tropical region, Malaysia. *Neural Comput. Appl.* 28, 893–905.
- Hart, J.D., Wehrly, T.E., 1986. Kernel regression estimation using repeated measurements data. *J. Amer. Statist. Assoc.* 81 (396), 1080–1088.
- He, X., Zhao, K., Chu, X., 2021. AutoML: A survey of the state-of-the-art. *Knowl.-Based Syst.* 212, 106622.
- Jamaludin, S.Z.M., Romli, N.A., Kasihmuddin, M.S.M., Baharum, A., Mansor, M.A., Marsani, M.F., 2022. Novel logic mining incorporating log linear approach. *J. King Saud Univ.-Comput. Inf. Sci.* 34 (10), 9011–9027.
- Karim, M.N., Munshi, S.U., Anwar, N., Alam, M.S., 2012. Climatic factors influencing dengue cases in dhaka city: a model for dengue prediction. *Indian J. Med. Res.* 136 (1), 32.
- Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J.C., Mathis, M., Brumby, S.P., 2021. Global land use/land cover with sentinel 2 and deep learning. In: *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. IEEE, pp. 4704–4707.
- Kasihmuddin, M.S.M., Jamaludin, S.Z.M., Mansor, M.A., Wahab, H.A., Ghadzi, S.M.S., 2022. Supervised learning perspective in logic mining. *Mathematics* 10 (6), 915.
- Kearns, M., Valiant, L., 1994. Cryptographic limitations on learning boolean formulae and finite automata. *J. ACM* 41 (1), 67–95.
- Lerner, D.N., Harris, B., 2009. The relationship between land use and groundwater resources and quality. *Land Use Policy* 26, S265–S273. <http://dx.doi.org/10.1016/j.landusepol.2009.09.005>, URL: <https://www.sciencedirect.com/science/article/pii/S0264837709001306>. *Land Use Futures*.
- Liu, K., Hu, X., Wei, Z., Li, Y., Jiang, Y., 2019. Modified Gaussian process regression models for cyclic capacity prediction of lithium-ion batteries. *IEEE Trans. Transp. Electr.* 5 (4), 1225–1236.
- Liu, F., Liao, S., Suykens, J., 2021. Kernel regression in high dimensions: Refined analysis beyond double descent. In: *International Conference on Artificial Intelligence and Statistics*. PMLR, pp. 649–657.
- Liu, Z., Pavao, A., Xu, Z., Escalera, S., Guyon, I., Junior, J.C.J., Madadi, M., Treguer, S., 2020. How far are we from true autotml: reflection from winning solutions and results of autotml challenge. In: *ICML Workshop*, Vol. 2020.
- Liuliakov, A., Hermes, L., Hammer, B., 2023. Autotml technologies for the identification of sparse classification and outlier detection models. *Appl. Soft Comput.* 133, 109942.
- Mack, Y.-p., Silverman, B.W., 1982. Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrscheinlichkeitstheor. Verwandte Geb.* 61, 405–415.

- Meza-Gastelum, M.A., Campos-Gaytán, J.R., Ramírez-Hernández, J., Herrera-Oliva, C.S., Villegas-León, J.J., Figueroa-Núñez, A., 2022. Review of groundwater withdrawal estimation methods. *Water* 14 (17), <http://dx.doi.org/10.3390/w14172762>, URL: <https://www.mdpi.com/2073-4441/14/17/2762>.
- Muhammad Sidik, S.S., Zamri, N.E., Mohd Kasihmuddin, M.S., Wahab, H.A., Guo, Y., Mansor, M.A., 2022. Non-systematic weighted satisfiability in discrete hopfield neural network using binary artificial bee colony optimization. *Mathematics* 10 (7), 1129.
- Nadiri, A.A., Naderi, K., Khatibi, R., Gharekhani, M., 2019. Modelling groundwater level variations by learning from multiple models using fuzzy logic. *Hydrol. Sci. J.* 64 (2), 210–226.
- Naem, M., Lee, D.C., 2011. Low-complexity joint transmit and receive antenna selection for MIMO systems. *Eng. Appl. Artif. Intell.* 24 (6), 1046–1051.
- Nagar, J., Chaturvedi, S.K., Soh, S., Singh, A., 2023. A machine learning approach to predict the k-coverage probability of wireless multihop networks considering boundary and shadowing effects. *Expert Syst. Appl.* 226, 120160.
- Nair, S.S., Sindhu, G., 2016. Groundwater level forecasting using artificial neural network. *Int. J. Sci. Res. Publ.* 6 (1), 2250–3153.
- Narasimhan, T.N., Witherspoon, P.A., 1982. Overview of the finite element method in groundwater hydrology. In: *Finite Elements in Water Resources: Proceedings of the 4th International Conference, Hannover, Germany, June 1982*. Springer, pp. 29–44.
- Niranjanika, M., Kumar, A., Beg, Z., Singh, A., Swarnkar, S., Gaurav, K., 2022. Groundwater variability in a semi-arid river basin, central India. *Hydrology* 9 (12), 222.
- Omar, P.J., Gaur, S., Dwivedi, S., Dikshit, P., 2019. Groundwater modelling using an analytic element method and finite difference method: an insight into lower ganga river basin. *J. Earth Syst. Sci.* 128, 1–10.
- Patel, S., Eldho, T., Rastogi, A.K., Rabinovich, A., 2022. Groundwater parameter estimation using multiquadric-based meshfree simulation with covariance matrix adaptation evolution strategy optimization for a regional aquifer system. *Hydrogeol. J.* 30 (7), 2205–2221.
- Patel, S., Rastogi, A., 2017. Meshfree multiquadric solution for real field large heterogeneous aquifer system. *Water Resour. Manage.* 31, 2869–2884.
- Perrone, M.P., Cooper, L.N., 1995. When networks disagree: Ensemble methods for hybrid neural networks. In: *How We Learn; how We Remember: Toward an Understanding of Brain and Neural Systems: Selected Papers of Leon N Cooper*. World Scientific, pp. 342–358.
- Pham, Q.B., Kumar, M., Di Nunno, F., Elbeltagi, A., Granata, F., Islam, A.R.M.T., Talukdar, S., Nguyen, X.C., Ahmed, A.N., Anh, D.T., 2022. Groundwater level prediction using machine learning algorithms in a drought-prone area. *Neural Comput. Appl.* 34 (13), 10751–10773.
- Poole, M.A., O'Farrell, P.N., 1971. The assumptions of the linear regression model. *Trans. Inst. Br. Geogr.* 145–158.
- Raj, R., Mathew, J., Kannath, S.K., Rajan, J., 2023. StrokeViT with automl for brain stroke classification. *Eng. Appl. Artif. Intell.* 119, 105772.
- Rasmussen, C.E., Williams, C.K., et al., 2006. *Gaussian Processes for Machine Learning*, Vol. 1. Springer.
- Raturi, A., Singh, H., Kumar, P., Chanda, A., Shukla, N., 2022. Characterizing the post-monsoon CO₂, CH₄, N₂O, and H₂O vapor fluxes from a tropical wetland in the Himalayan foothill. *Environ. Monit. Assess.* 194 (2), 50.
- Ravenscroft, P., Lytton, L., 2022. *Seeing the Invisible: A Strategic Report on Groundwater Quality*. Technical Report, pp. 56–65, URL: <https://openknowledge.worldbank.org/handle/10986/37197>.
- Sahin, E.K., Demir, S., 2023. Greedy-automl: A novel greedy-based stacking ensemble learning framework for assessing soil liquefaction potential. *Eng. Appl. Artif. Intell.* 119, 105732.
- Shlien, S., 1990. Multiple binary decision tree classifiers. *Pattern Recognit.* 23 (7), 757–763.
- Singh, A., Amutha, J., Nagar, J., Sharma, S., 2023a. A deep learning approach to predict the number of k-barriers for intrusion detection over a circular region using wireless sensor networks. *Expert Syst. Appl.* 211, 118588.
- Singh, A., Amutha, J., Nagar, J., Sharma, S., Lee, C.-C., 2022a. Automl-ID: Automated machine learning model for intrusion detection using wireless sensor network. *Sci. Rep.* 12 (1), 9074.
- Singh, A., Amutha, J., Nagar, J., Sharma, S., Lee, C.-C., 2022b. Lt-fs-id: Log-transformed feature learning and feature-scaling-based machine learning algorithms to predict the k-barriers for intrusion detection using wireless sensor network. *Sensors* 22 (3), 1070.
- Singh, A., Gaurav, K., 2023. Deep learning and data fusion to estimate surface soil moisture from multi-sensor satellite images. *Sci. Rep.* 13 (1), 2251.
- Singh, A., Gaurav, K., Rai, A.K., Beg, Z., 2021a. Machine learning to estimate surface roughness from satellite images. *Remote Sens.* 13 (19), 3794.
- Singh, A., Gaurav, K., Sonkar, G.K., Lee, C.-C., 2023b. Strategies to measure soil moisture using traditional methods, automated sensors, remote sensing, and machine learning techniques: review, bibliometric analysis, applications, research findings, and future directions. *IEEE Access*.
- Singh, A., Kotiyal, V., Sharma, S., Nagar, J., Lee, C.-C., 2020. A machine learning approach to predict the average localization error with applications to wireless sensor networks. *IEEE Access* 8, 208253–208263.
- Singh, A., Mehra, M., Kumar, A., Niranjanika, M., Priya, D., Gaurav, K., 2023c. Leveraging hybrid machine learning and data fusion for accurate mapping of malaria cases using meteorological variables in western India. *Intell. Syst. Appl.* 17, 200164.
- Singh, A., Nagar, J., Sharma, S., Kotiyal, V., 2021b. A Gaussian process regression approach to predict the k-barrier coverage probability for intrusion detection in wireless sensor networks. *Expert Syst. Appl.* 172, 114603.
- Svozil, D., Kvasnicka, V., Pospichal, J., 1997. Introduction to multi-layer feed-forward neural networks. *Chemometr. Intell. Lab. Syst. Syst.* 39 (1), 43–62.
- Szidarovszky, F., Coppola, Jr., E.A., Long, J., Hall, A.D., Poulton, M.M., 2007. A hybrid artificial neural network-numerical model for ground water problems. *Groundwater* 45 (5), 590–600.
- Tao, H., Hameed, M.M., Marhoon, H.A., Zounemat-Kermani, M., Salim, H., Sungwon, K., Sulaiman, S.O., Tan, M.L., Sa'adi, Z., Mehr, A.D., et al., 2022. Groundwater level prediction using machine learning models: A comprehensive review. *Neurocomputing*.
- Van Eck, N.J., Waltman, L., 2011. Vosviewer manual. Manual for VOSviewer version 1.
- Vapnik, V., Golowich, S., Smola, A., 1996. Support vector method for function approximation, regression estimation and signal processing. In: *Advances in Neural Information Processing Systems*, Vol. 9.
- Wang, M., Cheng, X., Zhu, X., 2014. Matching pursuit-based singular vectors estimation for large MIMO beamforming. *Electron. Lett.* 51 (1), 56–57.
- Wen, X., Feng, Q., Yu, H., Wu, J., Si, J., Chang, Z., Xi, H., 2015. Wavelet and adaptive neuro-fuzzy inference system conjunction model for groundwater level predicting in a coastal aquifer. *Neural Comput. Appl.* 26, 1203–1215.
- Wu, C., Yu, Q., Yi, K., 2012. Least-squares support vector machine-based learning and decision making in cognitive radios. *IET Commun.* 6 (17), 2855–2863.
- Wunsch, A., Liesch, T., Broda, S., 2021. Groundwater level forecasting with artificial neural networks: a comparison of long short-term memory (LSTM), convolutional neural networks (CNNs), and non-linear autoregressive networks with exogenous input (NARX). *Hydrol. Earth Syst. Sci.* 25 (3), 1671–1687.
- Yaseen, Z.M., El-Shafie, A., Afan, H.A., Hameed, M., Mohtar, W.H.M.W., Hussain, A., 2016. RBFNN versus FFNN for daily river flow forecasting at johor river, Malaysia. *Neural Comput. Appl.* 27, 1533–1542.
- Zamri, N.E., Azhar, S.A., Sidik, S.S.M., Mansor, M.A., Kasihmuddin, M.S.M., Pakrudin, S.P.A., Pauzi, N.A., Nawi, S.N.M., 2022. Multi-discrete genetic algorithm in hopfield neural network with weighted random k satisfiability. *Neural Comput. Appl.* 34 (21), 19283–19311.
- Zhongming, Z., Linong, L., Xiaona, Y., Wei, L., et al., 2021. UN world water development report 2021 'valuing water'.
- Zhou, Z., Hooker, G., 2021. Unbiased measurement of feature importance in tree-based methods. *ACM Trans. Knowl. Discov. Data (TKDD)* 15 (2), 1–21.