ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

A New Method for Determination of Optimal Borehole Drilling Location Considering Drilling Cost Minimization and Sustainable Groundwater Management

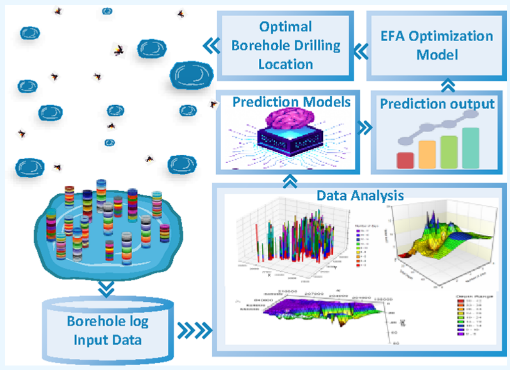
Anam Nawaz Khan, Bong Wan Kim, Atif Rizwan, Rashid Ahmad, Naeem Iqbal, Kwangsoo Kim, and Do Hyeun Kim**

Cite This: *ACS Omega* 2023, 8, 10806−10821

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Drilling boreholes for the exploration of groundwater incurs high cost with potential risk of failures. However, borehole drilling should only be done in regions with a high probability of faster and easier access to water-bearing strata, so that groundwater resources can be effectively managed. However, regional strati-graphic uncertainties drive the decision of the optimal drilling location search. Unfortunately, due to the unavailability of a robust solution, most contemporary solutions rely on physical testing methods that are resource intensive. In this regard, a pilot study is conducted to determine the optimal borehole drilling location using a predictive optimization technique that takes strati-graphic uncertainties into account. The study is conducted in a localized region of the Republic of Korea using a real borehole data set. In this study we proposed an enhanced Firefly optimization algorithm based on an inertia weight approach to find an optimal location. The results of the classification and prediction model serve as an input to the optimization model to implement a well-crafted objective function. For predictive modeling a deep learning based chained multioutput prediction model is developed to predict groundwater-level and drilling depth. For classification of soil color and land-layer a weighted voting ensemble classification model based on Support Vector Machines, Gaussian Naïve Bayes, Random Forest, and Gradient Boosted Machine is developed. For weighted voting, an optimal set of weights is determined using a novel hybrid optimization algorithm. Experimental results validate the effectiveness of the proposed strategy. The proposed classification model achieved an accuracy of 93.45% and 95.34% for soil-color and land-layer, respectively. While the mean absolute error achieved by proposed prediction model for groundwater level and drilling depth is 2.89% and 3.11%, respectively. It is found that the proposed predictive optimization framework can adaptively determine the optimal borehole drilling locations for high strati-graphic uncertainty regions. The findings of the proposed study provide an opportunity to the drilling industry and groundwater boards to achieve sustainable resource management and optimal drilling performance.



■ INTRODUCTION

Drilling boreholes for groundwater exploration and acquisition incurs high costs bearing significant risks and failures; therefore, the need for finding optimal borehole drilling locations cannot be overemphasized. Due to increased population, agriculture, industry, and urbanization, the rate at which groundwater is being pumped out has increased dramatically. The reliance on groundwater by around a third of the world's population has contributed to its gradual depletion. Nevertheless, groundwater is not an infinite resource, recognizing its potential is critical to ensuring its future usage. For these reasons drilling companies require a robust solution for finding the best borehole drilling location to extract water. Researchers demonstrate that[1] inappropriate selection and placement of boreholes account for a billion dollar increase in allocated budgets. Due to high cost many drilling companies undergo a third-rate investigation for finding an appropriate drilling location, resulting in million-dollar resource losses to companies per annum. Drilling engineers commonly conduct site investigation data and use prior knowledge of local geology for an optimal drilling location search; however, this method might face serious difficulties in cases when there is a lack of sufficient data.[2] Moreover it is challenging to directly examine, characterize, or measure the subsurface environment due to its heterogeneity and complex-

Received: October 26, 2022
Accepted: February 1, 2023
Published: March 16, 2023

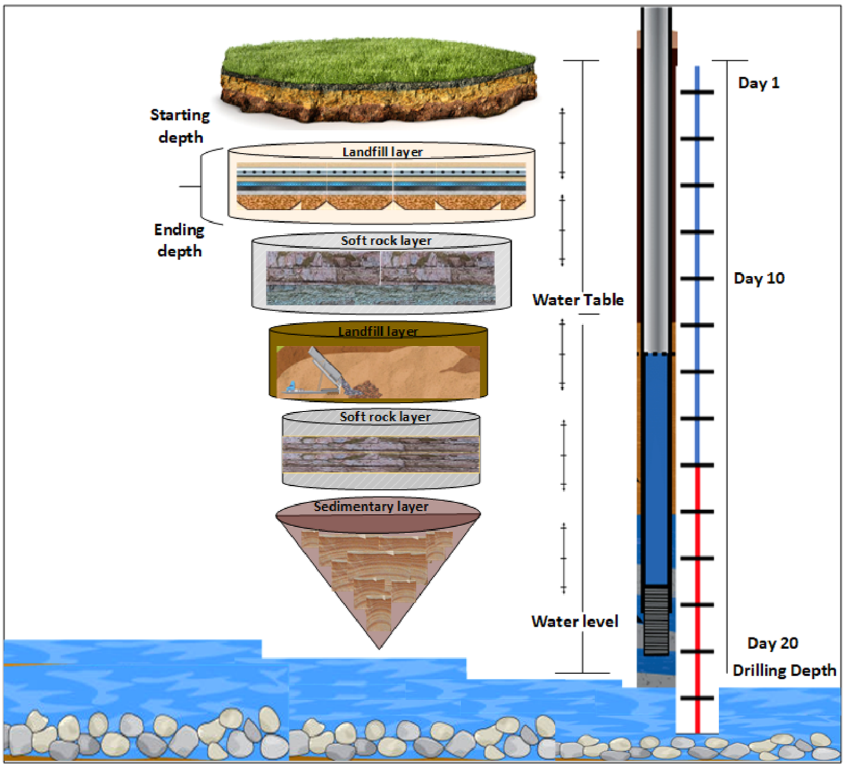


Figure 1. Overview of subsurface environment for groundwater extraction.

ity. Subsurface heterogeneity can have a significant influence on groundwater acquisition. Subsurface heterogeneity refers to variations in the hydro-geological properties of subsurface materials, such as differences in rock type, porosity, and permeability. The geological conditions and the formation structures have a significant impact on the water-storing capacity of a rock. For instance in the case of granite, a crystalline rock, there is almost no room for water storage. River gravel and limestone, on the other hand, are examples of loose soils that can retain and release significant volumes of groundwater. The permeability of a formation determines how quickly water can move through it, and this in turn is affected by the pore and void sizes and the connectivity between them. Additionally topography, climate variability, land use, vegetation parameters, hydrology, and human activities are just few of the many factors that can influence groundwater. In general, subsurface heterogeneity can affect groundwater acquisition in the following ways:

(1) Influencing the location of groundwater. Subsurface heterogeneity can cause variations in the location of groundwater, with some areas having higher or lower levels of groundwater than others. This can make it difficult to predict the location of groundwater resources and can affect the feasibility of drilling a borehole in a particular location.

(2) Affecting the quantity of groundwater. Subsurface heterogeneity can influence the amount of groundwater that is available in a particular area. In some cases, subsurface heterogeneity may cause variations in the porosity and permeability of subsurface materials, leading to differences in the amount of groundwater that can be stored and extracted. The potential of the groundwater is evaluated based on the geological and hydro-geological circumstances, and in accordance the required drilling resources, suitable rig types and budgets are suggested. Overall, subsurface heterogeneity can have a significant influence on groundwater acquisition, and it is important to consider these factors when selecting a location for a borehole and assessing the potential for groundwater extraction.

There are a number of approaches for the determination of drilling location. Recently geographical information systems (GIS) and remote sensing methods have become important tools for optimal drilling sites for exploration and management of groundwater resources. The adaption of a GIS based solution is attributed to their ability to analyze massive volumes of spatial data. Besides physical and nonphysical testing methods, the research on groundwater extraction and optimal bore-well placement is mainly divided into two categories, data-driven[3] and knowledge-based.[4] The former method is based on historical data, while the latter relies on expert opinions. Statistical models and data-mining based approaches are subtypes of data-driven methods. Traditional statistical methods and machine learning methods often employ bivariate and multivariate methods.[5] In recent years, numerous data mining techniques have been successfully applied to address complex real-world problems. Finding an optimal drilling site through machine learning methods brings considerable financial gains by accurately modeling down-hole conditions.[6] Example solutions for modeling groundwater and hydro-geological problems include Support Vector Machine,[7] Adaptive Neuro-Fuzzy Inference System (ANFIS),[8] the Artificial Neural Network (ANN),[9] Boosted Regression Tree (BRT),[10] Adaptive Boosting (Ada-Boost), Naive Bayes (NB),[11] and the Convolution Neural Network (CNN).[12] In addition, they find extensive application in sectors like land, flood, and soil erosion. Likewise, their adaption in geo-engineering problems have proven to be a substantial advantage. Meta-heuristic optimization has been recently adapted by scientists for use in hydro-geological contexts.[13] In order to find optimal parameters, certain meta-

https://doi.org/10.1021/acsomega.2c06854
ACS Omega 2023, 8, 10806−10821

heuristic algorithms are employed to tune parameters of machine learning models.

Unfortunately, the scope of existing studies is limited because of model parameters and data set accuracy. While manually tuning model parameters and hyper-parameters or using default values often results into poor generalization ability, the appropriate selection of model parameters and relevant features is critical for borehole location. Both of these criteria influence the dependability and precision of practical models. Till now there is no quantitative or objective method for selecting the right locations taking into account strati-graphic uncertainty. Clearly, there is a dire need to devise a mechanism for finding an optimal drilling location for groundwater extraction.[14]

2. RELATED WORK

This section presents an overview of existing solutions for determination of optimal borehole locations for groundwater extraction and potential mapping. Finding an optimal borehole location is highly desirable to meet the diversified needs of safe drinking water and productive drilling operations.[15] Knowledge about regional subsurface features, such as the spatial variability of soil at a given borehole location, is crucial for the selection of the drilling site.[16]

The down-hole environment can have a significant influence on the success of bore-well drilling operations. High temperatures and pressures can cause damage to drilling equipment and decrease its efficiency, while complex geology can make it difficult to navigate the well bore and maintain a stable borehole. Additionally, the down-hole environment can affect the choice and performance of drilling fluids, which are used to cool and lubricate the drill bit and remove cuttings from the well bore. Careful planning and attention to the down-hole environment can help ensure the success of bore-well drilling operations. Figure 1 illustrates the down-hole environment for acquisition of groundwater. The figure presents an overview of geological conditions within a well bore. A down-hole environment refers to the conditions and surroundings within a borehole or well. It is a challenging environment to work in, as it can be difficult to access and operate equipment at such depths, and the conditions can be harsh and unpredictable.

Groundwater is a scarce resource that plays a vital role in fostering economic development. In addition to providing irrigation for over 278.8 million acres of agriculture, groundwater supplies drinking water to over two billion people.[17] Contemporary methods for determining the optimal borehole location rely on a predictive modeling of hydro-geological features.[18] Predictive analytics facilitate the drilling companies in making the right decisions.[19] One of the most successful ways for managing groundwater is determination of accurate groundwater levels. In this context accurate mapping of the groundwater level is essential for water resource management. Groundwater level prediction models can help monitor the water level and make informed decisions. The knowledge enables the drilling companies to discover and mitigate issues such as water outages, pump failure, and factors impacting the geographic distribution of groundwater. Machine Learning techniques have gained enormous attention in the recent past[20] owing to their applicability across vast domains. For instance,[21] modeled groundwater using ANN, K-nearest neighbors and Classification and Regression trees (CART). Article[22] developed a CNN-LSTM model for the predicting groundwater level. Similarly, in ref 23 the drilling rate of penetration (ROP) is predicted using ANN. Ensemble learning techniques are a type of machine learning in which multiple models are trained to solve the same problem and their predictions are combined to produce a more accurate solution. Ensemble learning techniques are also widely adopted across wide application areas.[24] Article 25 proposed an optimized NN based solution for predicting circulation loss. The performance of ensemble models can be further improved using an optimal weight assignment scheme. To achieve this objective, ref 26 proposed a Fuzzy rule-based weight assignment scheme. while some of the studies reported the use of Meta-heuristics and probabilistic modeling frameworks for assigning weights to base models. The research work[20] developed a weighted voting ensemble classification model using Differential Evolution (DE) to categorize land layer and soil color. Meta-heuristics are algorithms that can be used to solve optimization problems, including the search for optimal borehole locations. These algorithms are designed to find approximate solutions to complex problems by exploring a large search space and identifying promising areas for further exploration. In the context of borehole location search, meta-heuristics can be used to identify locations that are most likely to yield a sufficient quantity of high-quality water. Some common meta heuristics for borehole location search include Genetic algorithms, Simulated Annealing, and Ant Colony Optimization. These algorithms can be customized to take into account a variety of factors, such as geology, hydrology, and drilling costs. Genetic Algorithm (GA) has been applied to find optimal testing locations for boreholes associated with pile foundation performance.[27] Several researchers focused on finding solutions for finding bore-well trajectories based on Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) algorithm.[28] Experimental results proved that meta-heuristics can achieve superior performance under operational constraints. Another work[29] proposed a biobjective GA for optimization of drilling location. A multiobjective optimization solution for bore-well placement is presented by ref 30. Similarly, a Greedy Randomized Adaptive Search Procedure (GRASP) for oil well drilling is proposed in ref 31. While ref 32 proposed a Nondominated Sorting Genetic Algorithm (NSGA-II) based solution for solving the multiobjective optimization problems. Literature reports the use of physical methods to determine the best drilling site for a borehole.[33] These methods involve collecting and analyzing data about the geology and hydrology of the area to identify locations that are most likely to yield a sufficient quantity of high-quality water.[34]

However, extensive literature review findings suggest that the data-driven methods provided comparatively better results in assessment of borehole and groundwater related parameters,[35] whereas conventional methods are unable to achieve the desired level of accuracy.[36] Despite all these efforts, the existing literature provides a little guidance on how to find optimal borehole drilling locations while taking into account the strati-graphic uncertainties.

Optimal borehole drilling for water extraction heavily depends on the strati-graphic properties of a drilling region. In order to take into account the uncertainties in the formation structure, it is important to conduct thorough research and analysis, using drilling logs and subsurface modeling to gain a better understanding of the conditions and potential risks at the site. To this aim we conducted a pilot study using land layer, soil composition, drilling depth, and groundwater level as key parameters for finding the best location for borehole drilling. The proposed study employs a real borehole data set comprised of borehole logging and lithology information. Our proposed work aims to find the best drilling location that helps with minimum drilling cost and computational complexity. The main contributions of the proposed work are concisely stated as follows: Development of a new predictive modeling assisted optimization strategy to find the best drilling location, that holds a rising groundwater level, minimum drilling depth, and softer and thicker land layer formations. To the best of our knowledge none of the existing solutions for optimal drilling site search have considered formation structures and hydro-geological attributes in the prediction assisted optimization framework. The proposed solution predicts the hydro-geological features and then finds the best locations using enhanced FA optimization model. The main contributions of the study are as follows:

1. Development of strati-graphic uncertainty aware predictive optimization strategy based on enhanced Firefly optimization model (EFA) to find optimal borehole drilling location to ensure efficient drilling operations.
2. Development of a chained multioutput prediction model based on long short-term memory (LSTM) to predict drilling depth and groundwater level.
3. Development of a soil color and land layer classification model using optimal weighted voting ensemble technique. A hybrid meta-heuristic Henry's Gas Solubility optimization algorithm and Brain-Storm optimization algorithm are employed to assign weights.
4. Borehole data analysis and feature engineering for extracting hidden insights from a data set.
5. To achieve faster convergence, we optimized to the performance of the original Firefly algorithm to mitigate the unpredictable behavior of fireflies when the mutual attraction between them is weak or absent.

3. MATERIAL AND METHODS

This section briefly describes the details of the experimental data set. Following that, a detailed data analysis and features engineering are presented. Lastly methodology of the proposed prediction and classification model is presented. The prediction model outputs drilling depth and groundwater level prediction results using a borehole log data set. While the weighted voting ensemble model provides the soil color and land layer classification results. The results of prediction model serve as an input to the optimization model to find the optimal borehole drilling location.

3.1. Borehole Log Data Set. The experiments are performed using real borehole data that are carefully examined for a valid transformation and modeling, and a diagnostic analysis is also provided to help unearth and extract important information from the data set. Based on data analysis, hidden insights are drawn from the data. The data set used in the proposed study is a real borehole log data set comprising 9287 instances of data representing 1987 unique boreholes drilled at several locations in a localized region of Republic of South Korea. The data set comprises geo-spatial and hydro-geological features. Table. 1 presents the details of the data set.

3.2. Spatial-Temporal Analysis of Borehole Logs. Figure 2 provides a visualization of borehole distribution with their respective location coordinates. The x-axis represents longitude values; whereas y-axis represents latitude values. It is evident from the figure that boreholes are not evenly distributed geographically. The concentration of boreholes is higher in areas where these factors are favorable for drilling, and lower in areas where they are less favorable.

Table 1. Detailed Feature Description

Feature	Description
Location coordinates (X, Y)	Drilling point location
Starting depth	Depth measured at the beginning
Ending depth	Depth measured at the end of drilling process
Altitude	The height of borehole location
Korean strata	Layer of stacked rocks/sediment
Groundwater level	Groundwater level water table information
Land layer	Groundwater is bounded above and below land
Soil color	layers (eight distinct classes)

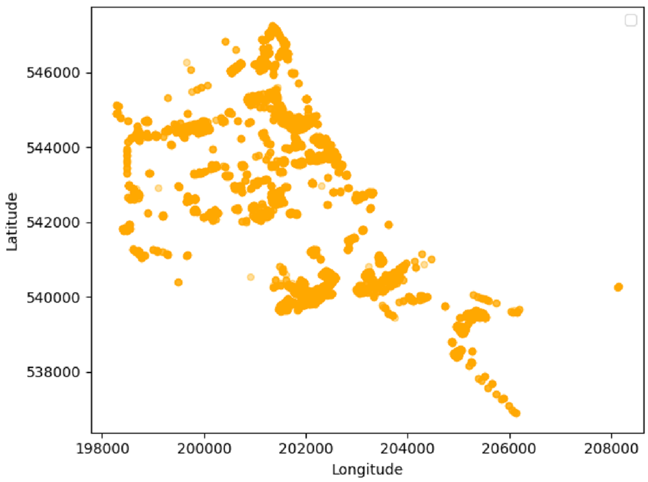


Figure 2. Distribution analysis of drilled borehole locations.

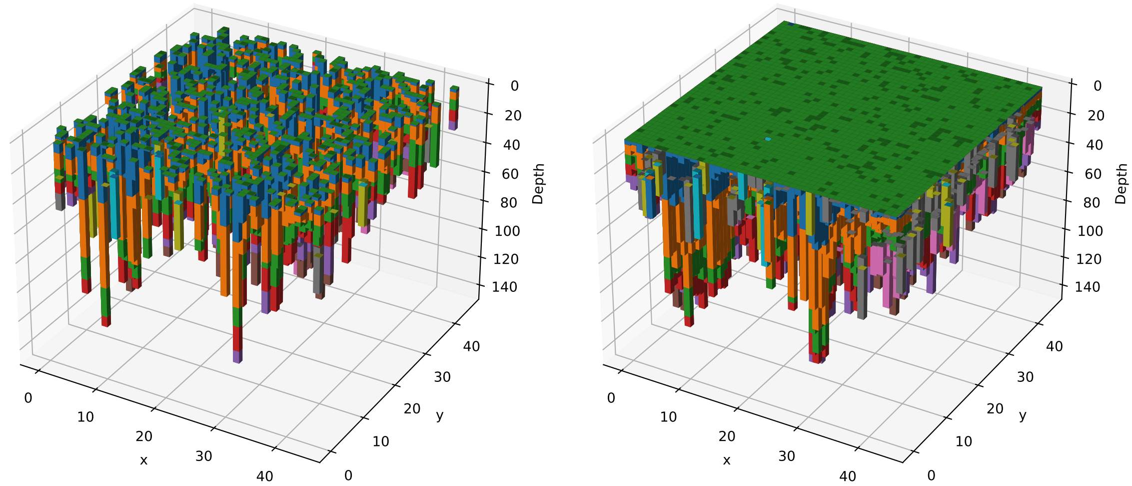
Figure 3(a) and 3(b) present the distribution of boreholes and their total depth in meters. Each vertical block shows one-day digging depth. The total height of the bar shows the total depth of a single bore in meters. The variation in drilling depth is associated with the water availability. A high groundwater table is a condition where the level of the groundwater is close to the surface of the ground so the drilling depth is minimum. This can happen in areas with high levels of precipitation or where the water table has risen due to an influx of water. On the other hand a low groundwater table is a condition where the level of the groundwater is significantly below the surface of the ground resulting in maximum drilling depths.

The proposed mechanism employs advanced feature engineering techniques to determine the underlying characteristics and unwind the peculiarities of formation structures against borehole drilling. As bore-hole drilling is a resource critical activity. Therefore, it is essential to calculate soil and land layer formation specific time consumed at each borehole site. For this average digging capacity for each soil color and land layer is computed. Moreover, total depth on each layer is also calculated based on starting depth and ending depth as follows:

$$\text{Depth}_t = \sum_{k=0}^{n} (D_e - D_s) \tag{1}$$

The total depth is the outcome of a difference function that calculates the difference of ending and starting depths. The hardness level of soil can be ascertain based on the statistics of average digging capacity, as average digging capacity is inversely proportional to the hardness level of land or rock layers and time and cost incurred. On account of this computing the average digging capacity is hugely substantiated. The average digging capacity is calculated as under

https://doi.org/10.1021/acsomega.2c06854
ACS Omega 2023, 8, 10806−10821



(a) Distribution of 500 boreholes (b) Entire distribution of boreholes

Figure 3. Visualization of boreholes in three-dimensional space.

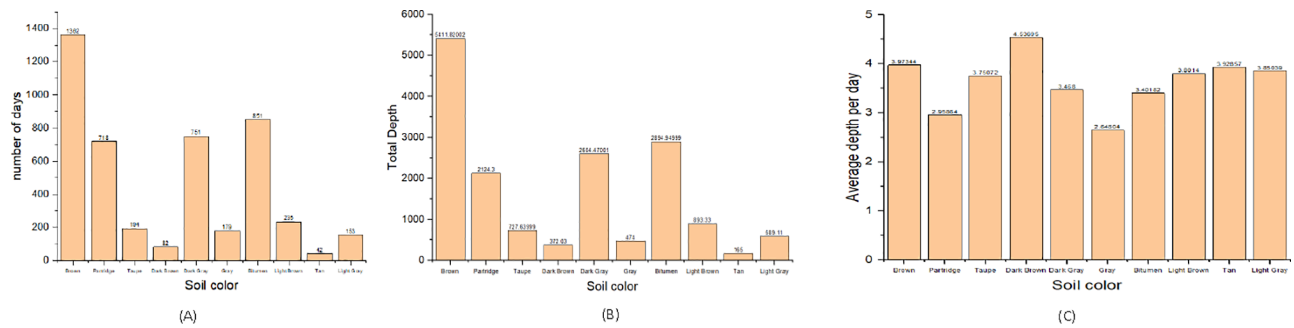


Figure 4. Spatiotemporal analysis of soil types.

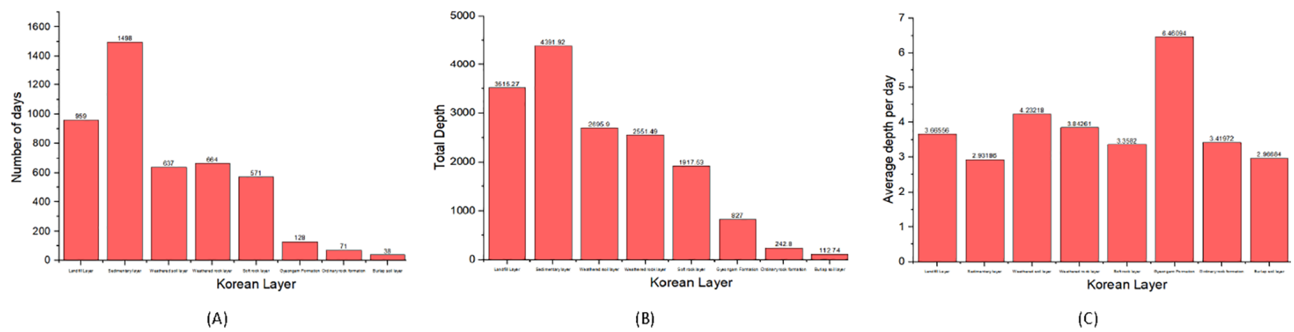


Figure 5. Spatiotemporal analysis of land layers.

$$\mathrm{DC}_{\mathrm{avg}}(S_{\mathrm{L}}) = \frac{\sum_{k=0}^{n} (D_{\mathrm{e}} - D_{\mathrm{s}})}{n} \tag{2}$$

where $\mathrm{DC}_{\mathrm{avg}}(S_{\mathrm{L}})$ is the average digging capacity of the land layer computed as difference of ending and starting depth divided by the number of instances for observed land layer.

Borehole data analysis is essential for successful drilling operations; thus, a curative borehole data analysis is done. As the subsurface conditions such as formation structures and lithology influence the drilling location selection process, examining data from previously drilled boreholes may be advantageous for reducing drilling costs and time. The borehole data include land layers information encountered during the borehole drilling process. These layers include landfill, sedimentary, weathered rock layer, burlap, alluvial, soft-rock, Gyeongam formation and remnant soil layers. Each layer possesses unique characteristics such as color, shape, hardness level, composition, and position. Figures 4 and 5 present a spatiotemporal analysis in terms of number of days spent on each distinct land layer and soil type, total depth achieved, and the average digging capacity. The analysis revealed that dark brown soil color has the highest per day average digging capacity of 4.3 m that translates into excessive rate of penetration due to its high compression ability. Comparatively, the Partridge soil color has the lowest average digging capacity of 2.3 m per-day, which means that this soil encounters more sand-stones with highly acidic pH levels. Figure 5 shows the spatial-temporal analysis of the land layer, the average digging capacity recorded across various land and rock

https://doi.org/10.1021/acsomega.2c06854
ACS Omega 2023, 8, 10806−10821

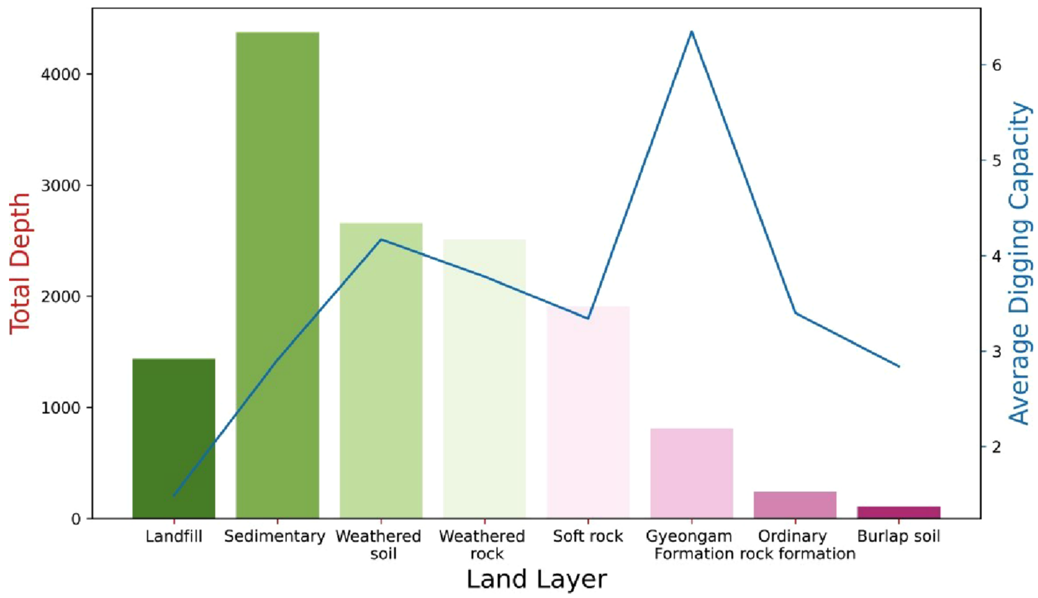


Figure 6. Analysis of land layer in terms of depth, hardness level, and average digging capacity.

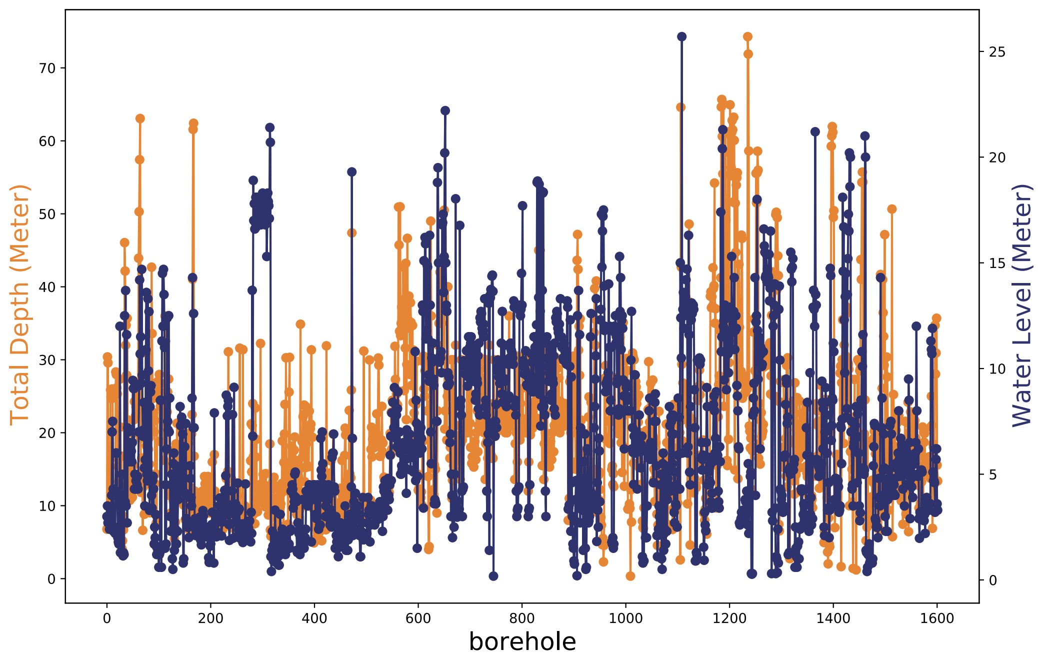


Figure 7. Analysis of water level and drilling depth.

strata. The Gyeongam formation has the highest average digging capacity, whereas the weathered rock layer attained second highest ADC. In contrast, the landfill layer has the lowest average digging capacity, indicating that it has a tougher composition and is susceptible to the risk of pipe clogging, resulting in operational failures and large resource losses. The research of land layers reveals that the sedimentary layer is frequently encountered during borehole drilling, while the weathered soil and rock layers rank second and third, respectively. Due to early disintegration characteristics, the burlap layer is the soil layer with the smallest total depth. The main components of a soft rock layer are silty clay and sand, while the lower layers are formed by weathered rocks. Gyeongam formation is a paddy layer type located in Gyeongam Province of South Korea. Among all, the sedimentary layer has a uniform lithology and texture with a composition of rounded grains containing many layers because of the disposition of sediments during their formation. Weathered rocks are known because of the higher porosity and lower bulk modulus, while burlap as a top soil layer is able to hold and keep soil in place. The alluvial soil layer is highly correlated to surface water presence composed of high absorption fine-grain fertile soil deposited by the flooded water usually dark in color carrying a mixture of clay slit and sand. The average digging capacity graph in Figure 6 demonstrate that the Gyeongam formation layer possesses the highest average digging capacity of 6.35 m due to low shear strength, while the landfill layer has the lowest average digging capacity of 1.49 m because of its stiff and hard nature.

For assessing the hydro-geological properties of subsurface land layers and subsurface structures at a particular drilling site,

https://doi.org/10.1021/acsomega.2c06854
ACS Omega 2023, 8, 10806−10821

the down-hole information obtained from borehole drilling is incredibly beneficial. The suggested work presented the relationship between total depth and water level through a time series analysis in Figure 7. The total depth of the borehole varies between 0 and 74.28 m. Due to the presence of variations in the formation structure, water table and drilling depth also varies. Moreover, the analysis made it clear that the majority of borehole sites found the groundwater at a water table depth of 10 m. Figure 8 shows the groundwater bearing strata across different geo-spatial locations. It is evident from the figures that groundwater level changes can vary greatly across different locations.

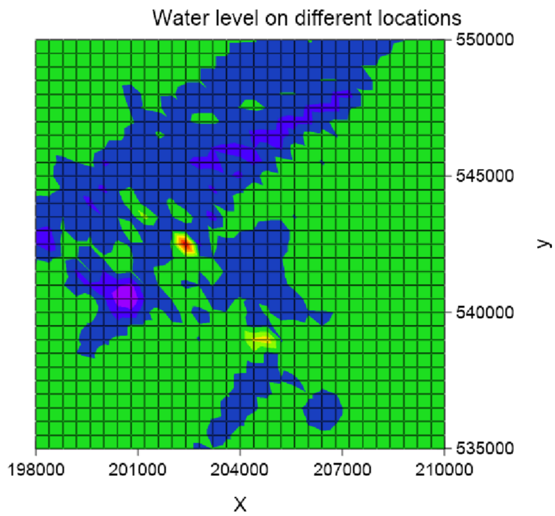


Figure 8. Analysis of groundwater bearing strata.

The geology of an area can affect the amount and distribution of groundwater, with some areas having more porous or permeable soils that allow for the infiltration of water, and other areas having less permeable soils that do not allow for the same level of infiltration. As a result, it is important to consider the specific conditions of a given location when assessing changes in groundwater levels.

3.3. Development of Regression and Classification Models. Predictive modeling aids effective drilling location selection decisions. The output of the prediction model directs the choice of the optimization model and is an integral component of the goal function designed to enhance the operational efficiency of drilling.

The optimization procedure employs key-parameters with related constraints, such as drilling depth, groundwater level, drilling region, and land layer formation in order to determine the optimal drilling site in terms of cost. The suggested solution assures that drilling companies and water boards maximize their productivity through optimal resource utilization and efficient drilling operations. Figure 9 explains the development of ensemble weighted voting classifier for classification of soil color and land-layer. Before further processing, the raw data are first preprocessed to remove any abnormalities.

The preprocessing steps involve removal of duplicate records, outlier detection, removal, and filtration of irrelevant attributes. Afterward the key features are extracted from the data to apply the process of feature engineering for construction of new features. Feature engineering refers to selection and transformation of relevant attributes for achieving superior performance of the predictive models by preparation of a best-fit data set for the learning algorithm. The new feature is then normalized using min−max scaler normalization. Following the preparation of the data, a comprehensive data analysis is conducted to provide valuable insight for the identification of patterns and trends and the creation of noteworthy conclusions. Afterward the data is split into train and test sets for the model training procedure.

In the case of ensemble learning frameworks, all the base models do not perform equally well. The proposed approach assigns weights to base learners based on their performance so that overall performance could be potentially improved. The developed ensemble model employs four base classifiers namely SVM, Gaussian NB, RF, and GBM with a varied set of hyper-parameter. The optimal set of weights is determined using a novel BHO optimization algorithm that is a hybrid of Henry's gas solubility optimization algorithm (HGSO) and Brain-storm Optimization algorithm (BSO). Figure 10 describes the flowchart of the proposed hybrid optimization algorithm. First, the base learners are trained using preprocessed data. Since the hyper-parameter plays a significant role in the model development process, Bayesian Optimization (BO) is used for optimal hyper-parameter tuning. The proposed linear SVM model utilizes a linear kernel function for classification of soil and land layer characteristics. The hyper-parameters set for the base model is merely a trade-off coefficient λ with a range of 1 and 2. Likewise, RF are trained with a hyper-parameter class weight value between 0 and 1, while GBM base learners are trained with a hyper-parameter step size value as 0.5 and 0.3, and maximum depth as 10 and 5, respectively. After model training the outcomes of SVM, RF, and GBM base learners are ensembled using weighted majority voting. A classification accuracy-based weight assignment policy is implemented. First, results from each base learner are obtained. Then, using the weights obtained during the training phase, a weighted majority vote of the results obtained from each base learner is used as a final prediction to classify soil color and land layer. The fitness function of the hybrid optimization algorithm is formulated as follows:

$$F(w) = \frac{\delta F_n + \gamma F_p}{T_p + T_n + \delta F_n + \gamma F_p} \quad (3)$$

The models that attain low classification accuracy are penalized by assigning less weight. To reduce a false negative rate initially the value of δ is set to less than γ, however; with increasing number of iterations the weight coefficients equalize to 1. While the value of δ and γ is computed as

$$\delta = 1 + \lambda\left[1 - \sin\left(\frac{\pi j}{2J}\right)\right] \quad (4)$$

$$\gamma = 1 - \lambda\left[1 - \sin\left(\frac{\pi j}{2J}\right)\right] \quad (5)$$

$$\lambda = \frac{\sigma(1 - 1/N)}{2} \quad (6)$$

Here "j" and "J" represent the current iteration and maximum iteration, respectively, while "N" denotes the population size. The flowchart of the HBSO is depicted in Figure 3. The algorithm begins with defining inputs such as size of population, upper and lower bounds of solution, dimension of problem, number of maximum iterations, and algorithm specific

https://doi.org/10.1021/acsomega.2c06854
ACS Omega 2023, 8, 10806−10821

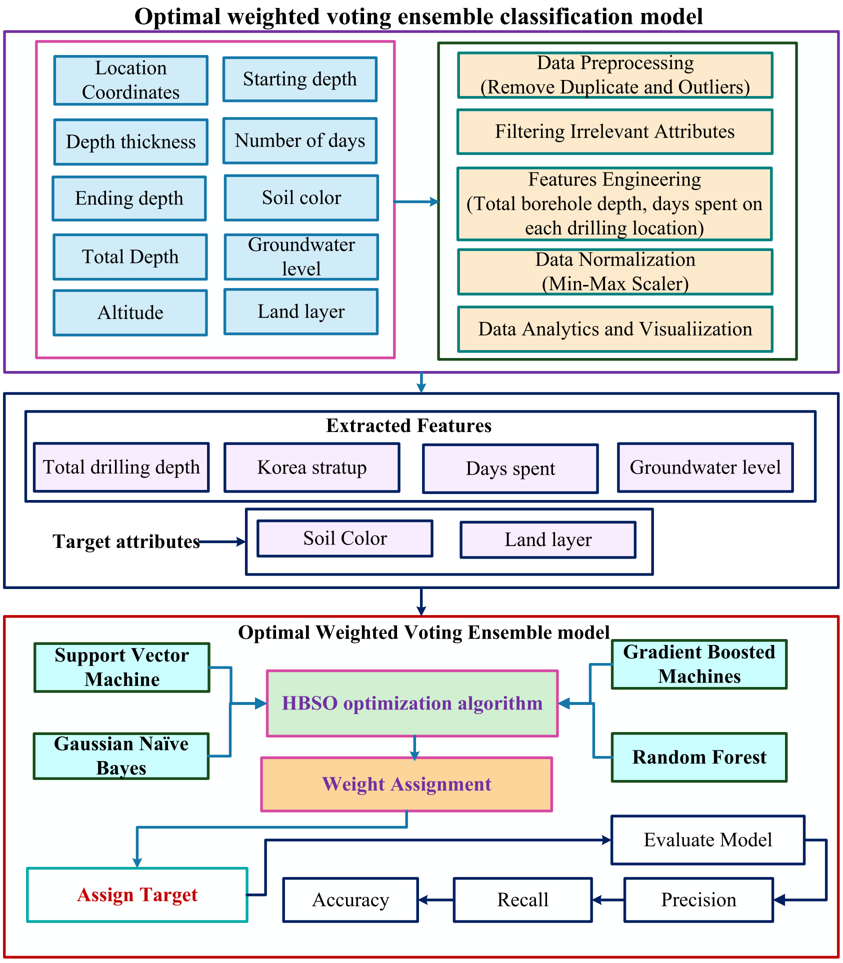


Figure 9. Operational overview of optimal weighted voting ensemble classifier.

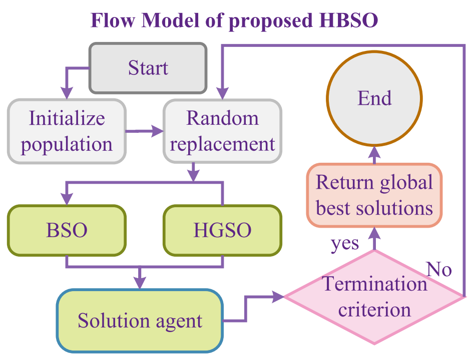


Figure 10. Flowchart of proposed optimal weighted voting mechanism

parameters. Afterward population is randomly initialized considering the upper and lower bounds of the solution space. The next stage, which is based on the greedy approach, involves replacing the best member of a group with a solution that has been generated at random in order to investigate other potential solutions. Subsequently two populations undergo BSO and HGSO simultaneously and "N" members of the population are selected out of both populations. To avoid becoming trapped in local optima, a solution agent is utilized to select the worst-performing solutions and replace them with randomly generated ones. In conclusion, the best performing solutions are returned as output if the termination requirement is satisfied. Following the receipt of results from the regression and classification models, those outputs are then fed into the optimization model in order to locate the most optimal location for drilling a borehole.

3.4. Long Short-Term Memory Chained Multi-output Regression Model.

We established and evaluated the LSTM based chained multioutput regression model developed to predict water level and drilling depth. Figure 11 shows the architecture of the proposed chained multioutput regression model to predict the water level and total drilling depth. A chained multioutput regression model is a type of machine learning model that is used for multioutput regression, which is a type of regression analysis where there are multiple target variables that need to be predicted. In the proposed chained multioutput regression model, the predictions for each target variable are made sequentially, with the predictions for one target variable serving as input for the next target variable. This approach can help to improve the overall prediction performance of the model by leveraging the relationships between the target variables. To train a chained multioutput regression model, we first split the data set into training and test sets. We then trained a separate regression model for each target variable, using the other target variables as input features. The predictions from each model would then be used as input features for the next model, forming a "chain" of predictions. The model will first make predictions for the first target variable using the input features, and then use those predictions as input features for the next target variable, and so on. This sequential approach allows

https://doi.org/10.1021/acsomega.2c06854
ACS Omega 2023, 8, 10806−10821

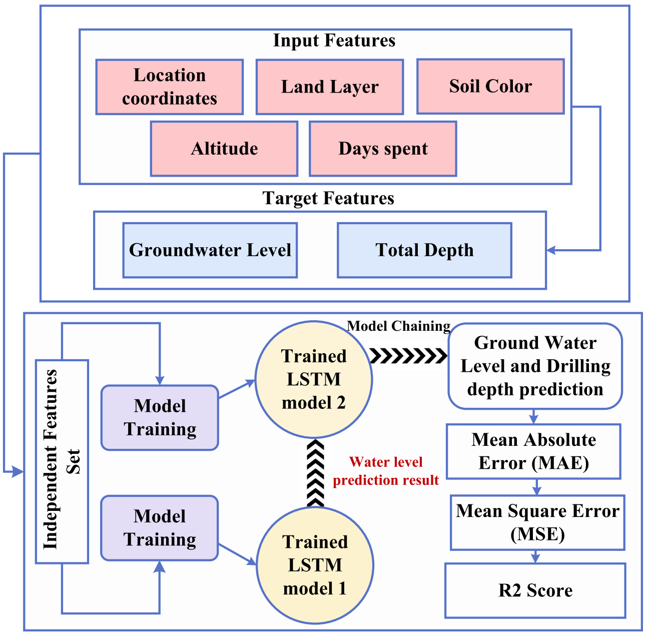


Figure 11. Architecture of proposed chained multioutput regression model.

the model to take into account the relationships between the target variables and make more accurate predictions.

4. PREDICTIVE OPTIMIZATION

The proposed prediction assisted optimization model finds the optimal drilling location by considering uncertainties in the formation structures. Formation structure refers to its physical characteristics, such as its composition, shape, and arrangement of its layers or strata. The formation structure can have a significant impact on the success of drilling operations, as it can affect the ease of drilling, the quality and quantity of the fluids that are produced, and the stability of the well bore. For locating an optimal borehole drilling site accurately, the drilling depth, water table, and land layer are predicted at the previously drilled borehole locations. To this aim an optimization model is developed. The optimization algorithm implements an objective function for drilling cost minimization. The goal of Firefly optimization is to find the solution that maximizes or minimizes the value of the objective function. Therefore, the design of the objective function is an important aspect of the Firefly optimization process. The objective function is carefully chosen to reflect the specific goals and constraints of the problem at hand. To solve this complex problem, we formulated a solution for the optimal drilling location by involving user preferences to devise a robust model that is scalable across dynamic environments and resource availability options. Since the allocated budgets and requirement of each borehole drilling project vary in terms of hydro-geological aspects, for the sake of simplicity we defined bounds for the selected parameters. For finding the optimal drilling location we considered three parameters: groundwater level, drilling depth, and land layer properties quantified in terms of hardness and softness level derived from average digging capacity. The optimal drilling location depends on the threshold values for the considered multiparameters. The threshold for each decision variables is defined as follows:

$$P_{\mathrm{WL,D,LL}} \in [\text{Low, Medium, High}] \tag{7}$$

The optimal drilling location finding mechanism outputs a location with minimum drilling cost based on the preset threshold defined for each constraint. In the proposed research study, the lower and upper bounds of the decision variables are quantified in terms of three point scale, i.e., low, medium, and high. In this way the drilling companies can opt for possible choices according to resource availability.

$$\begin{aligned} &P_{\mathrm{WL}}(\text{low}) = [<n_1, n_2>] \\ &P_{\mathrm{WL}}(\text{med}) = [<n_1, n_2>] \\ &P_{\mathrm{WL}}(\text{high}) = [<n_1, n_2>] \end{aligned} \tag{8}$$

The above equation shows the decision preference threshold for groundwater level. The available groundwater level preference options are medium (moderate), low (shallow) and high, to fulfill varied water level requirements.

$$\begin{aligned} &P_{\mathrm{D}}(\text{low}) = [<n_1, n_2>] \\ &P_{\mathrm{D}}(\text{med}) = [<n_1, n_2>] \\ &P_{\mathrm{D}}(\text{high}) = [<n_1, n_2>] \end{aligned} \tag{9}$$

Similarly for the land layer, any of the preferred choices are made in terms of hardness level or softness level while for drilling depth user preferred depth is selected. The values of individual parameters are restricted between $< n_1, n_2>$.

$$\begin{aligned} &P_{\mathrm{LL}}(\text{low}) = [<n_1, n_2>] \\ &P_{\mathrm{LL}}(\text{med}) = [<n_1, n_2>] \\ &P_{\mathrm{LL}}(\text{high}) = [<n_1, n_2>] \end{aligned} \tag{10}$$

The developed objective function focuses minimizing the drilling cost and time through a set of multiparameters and constraints. Each choice has an associated cost. Based on the cost less or more weights are assigned to each parameter for achieving the desired impact in optimal location search. The minimization objective function is defined as follows:

$$\mathrm{Obj}_{dc} = \mathrm{Min} \sum_{R=0}^{n} \left(\sum_{T \in (\pi L, D, LL)} (w_T \lambda_{T_n}) \right) \tag{11}$$

$$\begin{aligned} &WL \in C_{\mathrm{WL}} \\ &D \in C_{\mathrm{D}} \\ &SL \in C_{\mathrm{SL}} \\ &C_{\mathrm{WL}} \in [P_{\mathrm{WL_L}}, P_{\mathrm{WL_H}}] \\ &C_{\mathrm{D}} \in [P_{\mathrm{D_L}}, P_{\mathrm{D_H}}] \\ &C_{\mathrm{LL}} \in [P_{\mathrm{LL_L}}, P_{\mathrm{LL_H}}] \end{aligned} \tag{12}$$

Equation 11 provides the drilling cost minimization objective function. Equations 8, 9, and 10 depict the parameter's association, and eq 12 provides the decision bounds of the water level, drilling depth, land layer hardness/softness level. The constraint ensures the parameter values must be between maximum and minimum (upper and lower) bounds. As drilling a borehole through hard rocks can be a challenging and time-consuming process, the difficulty of the process depends on the type and strength of the rocks, as well as the depth of the borehole and the equipment being used. In general, drilling through hard rocks requires specialized equipment and

https://doi.org/10.1021/acsomega.2c06854
ACS Omega 2023, 8, 10806−10821

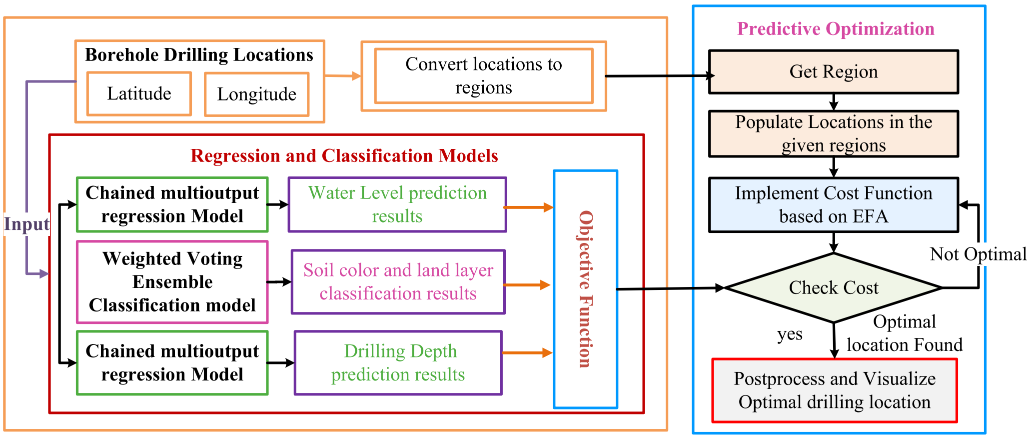


Figure 12. Optimal borehole drilling location search mechanism based on EFA.

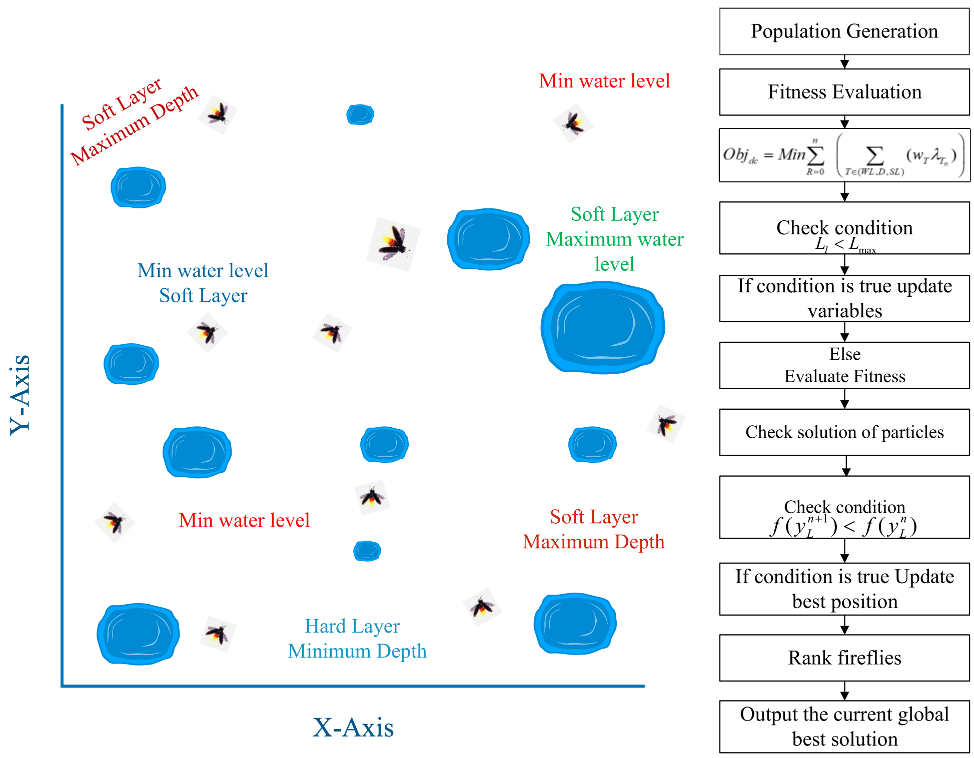


Figure 13. Flow-chart of the EFA optimization model.

experienced operators, and can take significantly longer than drilling through softer materials. For instance in the case of soft land layer, a shallow well can be dug with an auger drill, but in the case of hard rocks, a deeper hole will likely need to be dug with an air core drill. Thus, drilling through extremely difficult land layers such as rocks, or dolomite formations, requires expertise, special equipment, licenses, and proper training, adding an up-charge to the job. Figure 12 presents the predictive optimization mechanism for locating the optimal drilling location. First, the location will be converted to regions based on coordinates information. The output of prediction and classification models serve as input to the optimization model. The optimization model considers a particular region and populates locations in the given regions. The objective function is implemented through the EFA Optimization Algorithm. Based on the objective function the optimization algorithm finds the global optimal location considering the cost and associated constraints.

4.2. EFA Optimization Model. To find an optimal drilling location a cost minimization objective function and defined constraints evolve to relate to independent vectors and possess bounds between minimum and maximum values. Figure 13 presents the functional diagram of the EFA model.

The constraints get the minimum values in the first iteration until it receives maximum values with subsequent increments. The process continues until an optimal borehole drilling location is found. The proposed optimization model is attributed to the unknown solution space for the optimal location problem. This method can be safely utilized because solution space is nonconvex in nature. To achieve superior optimization, an EFA is developed. The algorithm implements a prevention strategy that restrains the random motion of fireflies in the case of zero or lowest possible attraction between fireflies, since the attraction between flies is affected by the value of the light absorption coefficient. The algorithm employs an inertia

https://doi.org/10.1021/acsomega.2c06854
ACS Omega 2023, 8, 10806−10821

Table 2. Comparative Analysis of the Proposed Optimal Weighted Voting Ensemble Model with Counterpart Solutions for Soil Color Classification

	Validation set			Test set		
Classifier	AC	P	R	AC	P	R
SVM	85.21 ± 1.22	84.21 ± 1.21	86.21 ± 1.68	87.32 ± 1.19	84.14 ± 0.79	86.74 ± 0.83
GBM	81.32 ± 0.67	77.55 ± 1.82	80.21 ± 1.16	84.56 ± 0.87	80.38 ± 1.01	79.91 ± 1.14
RF	78.01 ± 0.20	72.86 ± 1.09	73.55 ± 1.54	81.07 ± 2.04	74.61 ± 1.29	73.74 ± 0.87
GNB	82.31 ± 1.06	78.73 ± 1.27	84.23 ± 1.98	85.28 ± 1.81	83.87 ± 1.45	85.33 ± 1.23
MVE	89.55 ± 1.27	88.41 ± 1.33	90.22 ± 0.52	91.83 ± 1.27	92.94 ± 0.63	91.58 ± 0.66
Proposed WVC	91.24 ± 1.88	90.30 ± 0.62	89.46 ± 1.24	93.45 ± 0.56	91.86 ± 0.78	93.11 ± 1.25

Table 3. Comparative Analysis of the Proposed Optimal Weighted Voting Ensemble Model for Land Layer Classification

	Validation set			Test set		
Classifier	AC	P	R	AC	P	R
SVM	86.21 ± 0.34	85.59 ± 1.12	82.21 ± 1.02	88.32 ± 0.14	85.14 ± 1.45	85.74 ± 0.83
GBM	79.32 ± 0.78	79.55 ± 0.65	78.21 ± 0.45	83.56 ± 1.82	82.38 ± 0.52	84.91 ± 1.71
RF	82.01 ± 1.20	79.86 ± 0.55	84.55 ± 0.87	85.07 ± 1.06	83.61 ± 0.67	78.74 ± 0.89
GNB	81.67 ± 1.05	86.73 ± 0.76	85.23 ± 1.65	82.88 ± 1.37	79.87 ± 0.93	90.95 ± 0.63
MVE	90.55 ± 1.53	91.41 ± 1.87	90.81 ± 0.23	93.44 ± 0.86	94.9 ± 1.03	92.58 ± 0.54
Proposed WVC	92.18 ± 0.94	92.47 ± 0.62	89.46 ± 1.04	95.34 ± 0.71	96.86 ± 1.56	95.11 ± 1.76

weight approach for improving the standard Firefly algorithm to achieve superior convergence and global optimization. A least attraction framework is introduced for preventing firefly from random motion and is described as an attraction between flies in terms of Cartesian distance.

$$\alpha_{lm}(s_{lm}) = \alpha_{min} + (\alpha_0 - \alpha_{min}) j^{\gamma s^2 m} \tag{13}$$

where $\alpha_0 = 1.0$, and the minimum attraction between them can be [0,1] The developed inertia weight strategy involves logarithmic decrement. At the start, the inertia weight strategy is linear in nature; hence, a less bright fly is attracted by a brighter one such that its early motion drives it toward global search. With increasing iteration, the fly moves closer to the optimal value being in a steady state for enhancing the local search ability, avoiding vertical oscillations, and overshooting the problem. The equation for logarithm inertia weight is

$$w_x = w_1 - (\log_{adj}) \times (w_1 - w_2) \times \log_x x \tag{14}$$

The initial weight and final inertia weights are represented by w_1 and w_2, respectively, $\log_{adj}$ is the adjustment factor of logarithm, whereas 'x' and 'X' show the current and maximum number of iterations. For further dealing with random turbulence resulting in a slow convergence, a step adjustment factor is introduced as follows:

$$k = \tau^{df} X(j^{(-x/X)}) \tag{15}$$

The search dimension of a firefly is given as $\tau^{df} \in [0, 1]$. With increasing search dimensions, the random step size is decreased. The adjustment factor is introduced to limit the change in random step size of firefly to achieve performance gains. Lastly the position of fireflies is updated on the basis of attraction between a highly brighter initialization fly toward a lower initialized brightness using the following equation:

$$y_l^{x+1} = w_x(y_l^x S_{lm}) \times (y_l^x - y_m^x) + \beta^* \varphi_{lm} \tag{16}$$

Algorithm 1 presents stepwise enhanced FA optimization model. FA has received enormous attention in the recent past because of its vast applications. The detailed pseudocode for the optimization process is depicted in Algorithm 1.

Algorithm 1: An enhanced Firefly optimization algorithm for finding optimal drilling location for groundwater extraction

Data: General Initialization $y_i = (L = 1, 2, 3, \ldots n)$
Result: Optimal borehole drilling location
initialization;
Substitute the position vector of fireflies← $f(\vec{y}_i)$
Evaluate fitness of fireflies $L_L = f(\vec{y}_i)$
while *stopping criteria not satisfied* **do**
  if $L_L < L_{max}$ **then**
    Update cartesian distance
    Update attraction by using Eq.(13)
    Update W_x *using eq* (14)
    Update step adjustment factor using eq. (15)
    Update the position of current fireflies using eq. (16)
  else
    Evaluate the fitness of fireflies
    Check solution of the particles
if $f(y_L^{n+1}) < f(y_L^n)$ **then**
  Update best position
else
  Rank the fireflies and find the current global best g

5. EXPERIMENTAL RESULTS AND DISCUSSION

For performance evaluation the classification model Accuracy, Precision, and Recall/sensitivity are used as evaluation metrics. The results of the model are provided in Tables 2 and 3. Accuracy is a measure of how often the model correctly predicts the outcome. Precision is a measure of how often the model's predictions are correct when they predict a positive outcome, while Recall is a measure of how often the model predicts a positive outcome when the true outcome is positive. To analyze the prediction performance of a model, these metrics are calculated using the predicted outcomes and the true outcomes from a test data set. After that we compared the values of these metrics for counterpart models to determine which model has

https://doi.org/10.1021/acsomega.2c06854
ACS Omega 2023, 8, 10806−10821

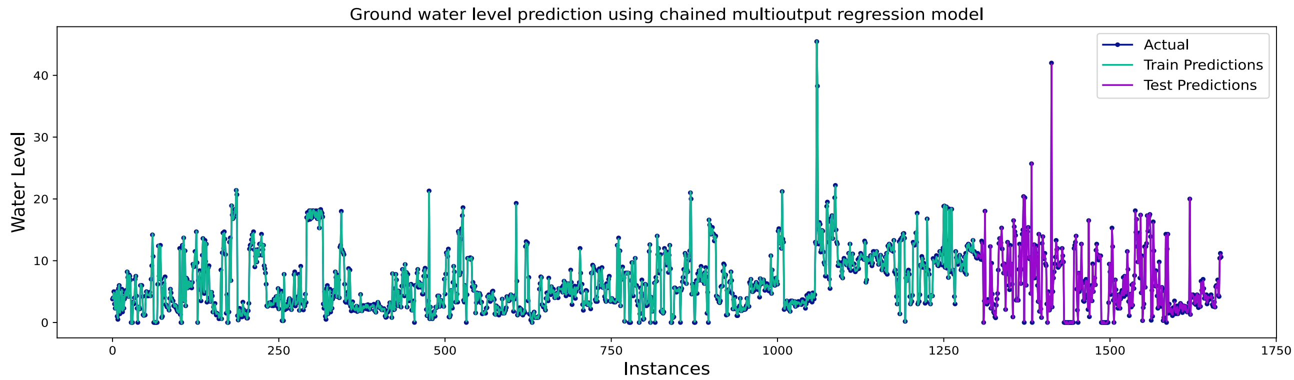


Figure 14. Ground water level prediction using chained multioutput regression model.

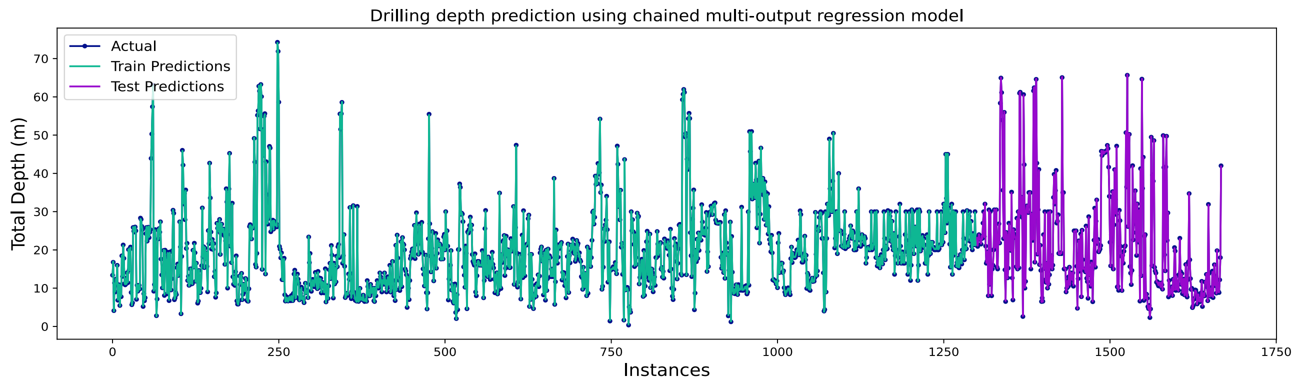


Figure 15. Drilling depth prediction using chained multioutput regression model.

the best prediction performance. For the training of weighted voting classifier, 10-fold validation has been employed. At each iteration nine out of 10-folds are used to train the model. The upper and lower bounds of standard deviation of the evaluation metrics Precision, Recall and Accuracy is reported for each classification model. The results of the proposed model are compared with standalone models as well as the majority voted ensemble model. The weights are assigned to each learner based on classification accuracy; thus, a candidate learner with higher accuracy is given more priority in decision making.

As target attributes, soil color and land layer have multiclasses so the classification results are based on a mean result of all classes. The comparative analysis of the proposed weighted voting ensemble classification model with baseline models was performed. Extensive experimentation revealed that the proposed ensemble model achieved higher accuracy and performance for both testing and validation set. The validation and testing accuracy achieved by the classification model for soil color classification is 91.24% and 93.4%, respectively, while the performance of the majority voting ensemble model is competitive to the proposed model with test accuracy of 91.83%. The rest of the stand-alone model failed to produce results with high accuracy.

For land layer classification the proposed weighted voting scheme attained 95.34% accuracy on test data. While the majority voting ensemble model achieved an accuracy of 93.44%. In the case of the independent SVM model, the accuracy dropped to 88.32%. The performance of the proposed ensemble model is attributed to the optimal weight assignment technique. The proposed approach optimally assigns weights to each base learner during the training phase to get results for each learning model that guides the final prediction model. Through incorporating the knowledge of all base learners, the proposed ensemble classification model outperformed the rest of the solutions.

The prediction results of groundwater level and drilling depth is presented in Figures 14 and 15. We employed the chained multioutput regression model for prediction and made a comparative analysis with conventional machine learning models. For evaluating the performance of regression models the mean absolute error (MAE), mean square error (MSE), and *R*-squared score (R2 score) have been employed. The mean absolute error (MAE) is a measure of the average magnitude of the error in a set of predictions. Mean squared error (MSE) is a measure of the average squared difference between the predicted values and the true values. The *R*-squared score, also known as the coefficient of determination, is a measure of the goodness of fit of a regression model. It is a statistic that provides a measure of how well the model fits the data. For model training, K-Fold cross-validation is employed. Closer inspection of actual training and testing prediction results reveals that the model produced remarkable results with lower error rate between actual and predicted values of groundwater level and drilling depth.

Table 4 compares the performance of proposed chained multioutput prediction model with conventional machine

Table 4. Performance Analysis of Chained Multi-Output Regression Model for Groundwater Level Prediction

Model	MAE	MSE	R2 score
Linear regression	7.34	122.56	64.14
K-nearest neighbor	3.65	60.71	79.02
Decision Tree	8.10	106.42	68.72
Proposed	2.89	53.73	85.43

https://doi.org/10.1021/acsomega.2c06854
ACS Omega 2023, 8, 10806−10821

learning models for groundwater level prediction. Accurate groundwater level prediction is an important factor to consider when choosing a drilling site. This is because the groundwater level can affect the success and sustainability of the drilling project. Therefore, accurately predicting the groundwater level at a potential drilling site can help ensure the success and sustainability of the project. This results are quite revealing that the proposed prediction model acquired the lowest MAE of 2.89%. While the Decision Tree model achieved a MAE score of 8.10%.

It is evident that compared to LSTM model the conventional machine learning models remained unable to generalize the data well, as the drilling depth varies across different geographical regions depending upon the availability of groundwater resource. Accurate drilling depth prediction is important for several reasons. In the context of borehole drilling, it can help to ensure that the borehole is drilled to the desired depth and that it reaches the target geological formation. For drilling depth prediction Table 5 shows that the proposed model achieved the lowest MAE score of 3.11% while the independent Decision Tree and K-nearest Neighbor classifier achieved MAE of 5.89% and 5.92%, respectively. The results of the proposed model are more precise than the counterpart conventional machine learning model. In summary the accurate prediction insights can help save resources by avoid overdrilling or underdrilling, which can be costly and time-consuming to correct. The performance evaluation and results suggest that the proposed chained multioutput regression model produced highly accurate results and performed fairly well in terms of MSE, MAE, and R2 score. The performance achieved by the proposed regression model is attributed to the ability of the LSTM model to map spatial and temporal correlations from data.

Table 5. Performance Analysis of Chained Multi-output Prediction Model for Drilling Depth Prediction

Model	MAE	MSE	R2 score
Linear regression	8.75	111.70	61.14
K-nearest neighbor	5.92	57.31	79.46
Decision Tree	5.89	110.36	75.83
Proposed	3.11	51.20	91.56

Table 5 presents the results of drilling depth. Accurate depth prediction of a drilling site can be an important factor in determining the optimal drilling site. Accordingly the proposed deep learning based prediction model achieved the lowest MAE score of 2.89%. Thus, the LSTM model outperformed all the counterpart solutions in terms of MAE, MSE, and R2 score. Lastly Figure 16 describes the simulation results of our proposed optimal drilling location finding mechanism along the complexity and convergence analysis of an enhanced Firefly metaheuristic algorithm that is inspired by the behavior of fireflies. The firefly algorithm uses the concept of attractiveness and light intensity to simulate the movement of fireflies in search of a mate. The algorithm works by having each firefly in the population move toward the firefly with the highest attractiveness. This movement continues until the fireflies converge on the firefly with the highest attractiveness, which represents the global optimum solution to the problem being solved.

Therefore, we first analyzed specific factors and constraints that are relevant to the optimal drilling location finding including geology of the area, the water table depth, availability of equipment and resources. Then we designed the Firefly optimization algorithm to take these factors into account and to search for solutions that meet the desired criteria. The optimization algorithm aim to find a location that provides the maximum water yield for a given drilling depth, or find the location that minimizes the overall cost of the drilling project. Figure 17 presents the results of optimal drilling point search using the EFA model. The results of the model prove the effectiveness of the developed solution based on low cost and high performing solutions. In terms of convergence, the firefly algorithm uses a combination of exploration and exploitation to find the global optimum solution. During the exploration phase, fireflies move randomly in search of new solutions. This helps to ensure that the algorithm does not get stuck in a local optimum. During the exploitation phase, fireflies move toward the firefly with the highest attractiveness, which helps to refine the solution and improve its quality. The result for the optimal drilling point search shows superior optimization results with a faster convergence. The algorithm has explored the search space and identified the locations that meet the specified constraints, providing a list of potential drilling location solutions last converging to an optimal location. The computational complexity analysis of the proposed EFA is $O(N)$ for the initialization phase while due to an inertia weight factor, the computational complexity for finding the optimal drilling point for the borehole placement is $O(N) + (3^{*}X^{*}N) + O(2^{*}P_{s}^{*}X^{*}\mathrm{obj}_{\mathrm{j_{dc}}})$. The ob-

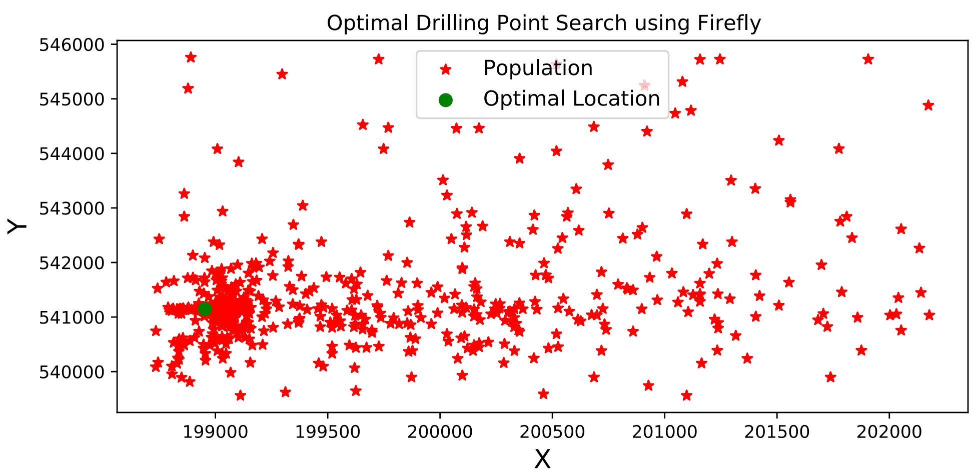


Figure 16. Optimal drilling location result by EFA optimization

https://doi.org/10.1021/acsomega.2c06854
ACS Omega 2023, 8, 10806−10821

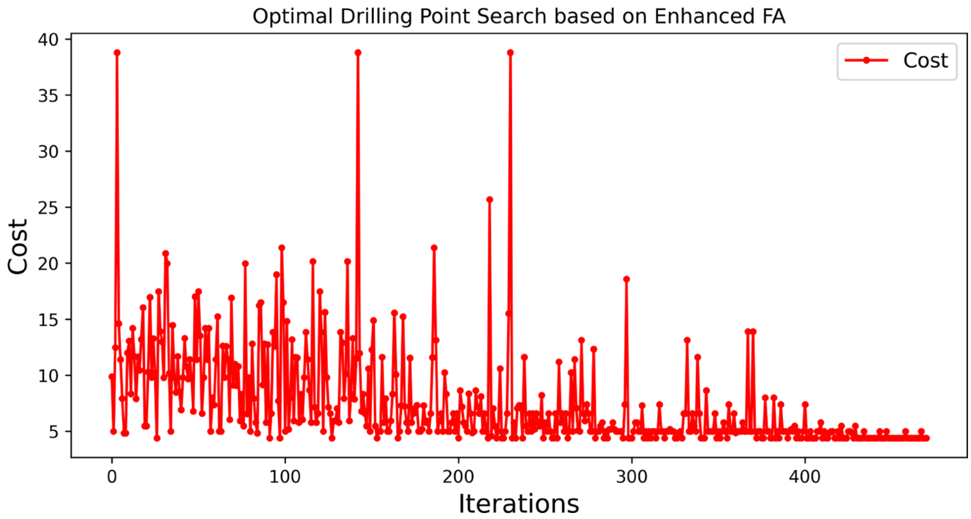


Figure 17. Convergence analysis of proposed EFA optimization algorithm.

jective function takes as input the values of the decision variables for a particular solution, and it returns a location that represents the quality of the solution.

CONCLUSION

In attempts to provide a reliable source of clean and safe water, the location of the borehole drilling is one of the many factors that becomes very crucial to take into consideration. However, optimal drilling location is highly dependent on hydrogeological properties of the site under consideration. To resolve this issue a predictive optimization approach is proposed considering three critical stratigraphic factors. For prediction a competitive chained multioutput prediction model based on the LSTM model is developed to predict the drilling depth and groundwater level simultaneously. Furthermore, a novel weighted voting ensemble classification model is developed for classification of soil color and land layer. The weighting mechanism is devised using a hybrid meta-heuristic optimization algorithm to assign weights to model by making use of extraordinary exploitation and exploration abilities with reduced computational cost. A specially designed fitness function is developed for assisting the classification model to achieve high accuracy. For finding the optimal location for bore well drilling, FA based on an inertia weight approach is developed. The enhanced FA algorithm relies on an autotuned logarithm inertia weight to attain a minimal attraction phenomenon. The use of a step adjustment factor has greatly enhanced the optimization performance by providing robust solutions. The statistical goodness of fit of prediction and classification models is confirmed by the results of the hypothesis testing. Experimental results indicate a high level of concordance between the actual and predicted values that translates in to high predictive performance. Furthermore, the cost and convergence analysis results prove the effectiveness of the proposed prediction assisted optimization mechanism. The results of the proposed pilot study can open avenues for drilling companies and water boards to apply such frameworks for determination of optimal drilling sites.

AUTHOR INFORMATION

Corresponding Author

Do Hyeun Kim − *Advanced Technology Research Institute, Jeju National University, Jeju 63243, Republic of Korea*; Email: kimdh@jejunu.ac.kr

Authors

Anam Nawaz Khan − *Department of Computer Engineering, Jeju National University, Jeju 63243, Republic of Korea*; orcid.org/0000-0001-6260-5820

Bong Wan Kim − *Electronics and Telecommunications Research Institute (ETRI), Daejeon 34129, Republic of Korea*

Atif Rizwan − *Department of Computer Engineering, Jeju National University, Jeju 63243, Republic of Korea*

Rashid Ahmad − *Bigdata Research Center, Jeju National University, Jeju 63243, Republic of Korea; Department of Computer Science, COMSATS University Islamabad, Attock 43600, Pakistan*

Naeem Iqbal − *Department of Computer Engineering, Jeju National University, Jeju 63243, Republic of Korea*

Kwangsoo Kim − *Electronics and Telecommunications Research Institute (ETRI), Daejeon 34129, Republic of Korea*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c06854

Notes

The authors declare no competing financial interest.
The source code for finding the optimal borehole drilling location is available on Github Ana119/Optimal-Drilling-Location

ACKNOWLEDGMENTS

This research was supported by Energy Cloud R&D Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT (2019M3F2A1073387), and this work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant 21DCRU-B158151-02). Any correspondence related to this paper should be addressed to Dohyeun Kim

■ REFERENCES

(1) Crisp, M.; Jaksa, M.; Kuo, Y. Effect of borehole location on pile performance. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards* **2022**, *16*, 267−282.

(2) Chowdhury, A.; Jha, M.; Chowdary, V.; Mal, B. Integrated remote sensing and GIS-based approach for assessing groundwater potential in West Medinipur district, West Bengal, India. *International Journal of Remote Sensing* **2009**, *30*, 231−250.

(3) Arabameri, A.; Lee, S.; Tiefenbacher, J. P.; Ngo, P. T. T. Novel Ensemble of MCDM-Artificial Intelligence Techniques for Groundwater-Potential Mapping in Arid and Semi-Arid Regions (Iran). *Remote Sensing* **2020**, *12*, 490. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

(4) Kaliraj, S.; Chandrasekar, N.; Magesh, N. S. Identification of potential groundwater recharge zones in Vaigai upper basin, Tamil Nadu, using GIS-based analytical hierarchical process (AHP) technique. *Arabian Journal of Geosciences* **2014**, *7*, 1385−1401.

(5) Naghibi, S. A.; Pourghasemi, H. R. A Comparative Assessment Between Three Machine Learning Models and Their Performance Comparison by Bivariate and Multivariate Statistical Methods in Groundwater Potential Mapping. *Water Resources Management* **2015**, *29*, 5217−5236.

(6) Crisp, M.; Jaksa, M.; Kuo, Y. L. *Influence of Site Investigation Borehole Pattern and Area on Pile Foundation Performance.* 12th ANZ Young Geotechnical Professionals Conference, Australia, 2018.

(7) Serapiao, A. B.; Tavares, R. M.; Mendes, J. R. P.; Guilherme, I. R. Classification of petroleum well drilling operations using Support Vector Machine (SVM). *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)* **2006**, 145−145, DOI: 10.1109/CIMCA.2005.1631427.

(8) Moghaddam, D. D.; Rahmati, O.; Panahi, M.; Tiefenbacher, J.; Darabi, H.; Haghizadeh, A.; Haghighi, A. T.; Nalivan, O. A.; Tien Bui, D. The effect of sample size on different machine learning models for groundwater potential mapping in mountain bedrock aquifers. *CATENA* **2020**, *187*, 104421.

(9) Maiti, S.; Krishna Tiwari, R.; Kümpel, H.-J. Neural network modelling and classification of lithofacies using well log data: A case study from KTB borehole site. *Geophysical Journal International* **2007**, *169*, 733−746.

(10) Naghibi, S.; Pourghasemi, H.; Dixon, B. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. *Environmental Monitoring and Assessment* **2016**, *188*, 188.

(11) Pham, B. T.; Jaafari, A.; Phong, T. V.; Mafi-Gholami, D.; Amiri, M.; Van Tao, N.; Duong, V.-H.; Prakash, I. Naïve Bayes ensemble models for groundwater potential mapping. *Ecological Informatics* **2021**, *64*, 101389.

(12) He, M.; Zhang, Z.; Ren, J.; Huan, J.; Li, G.; Chen, Y.; Li, N. Deep convolutional neural network for fast determination of the rock strength parameters using drilling data. *International Journal of Rock Mechanics and Mining Sciences* **2019**, *123*, 104084.

(13) Khosravi, K.; Panahi, M.; Tien Bui, D. Spatial prediction of groundwater spring potential mapping based on an adaptive neuro-fuzzy inference system and metaheuristic optimization. *Hydrology and Earth System Sciences* **2018**, *22*, 4771−4792. Publisher: Copernicus GmbH.

(14) Bowles, J. E. *Foundation analysis and design*; McGraw-Hill, 1988.

(15) Liu, H.; Gjersvik, T. B.; Faanes, A. Subsea field layout optimization (part III) — the location-allocation problem of drilling sites. *J. Pet. Sci. Eng.* **2022**, *208*, 109336.

(16) Hassanipak, A.; Sharafodin, M. GET: a function for preferential site selection of additional borehole drilling. *Exploration and Mining Geology* **2004**, *13*, 139−146.

(17) Seidler, R.; Padalkina, K.; Buecker, H. M.; Ebigbo, A.; Herty, M.; Marquart, G.; Niederau, J. Optimal experimental design for reservoir property estimates in geothermal exploration. *Computational Geosciences* **2016**, *20*, 375−383.

(18) Iqbal, N.; Khan, A.; Rizwan, A.; Ahmad, R.; Kim, B. W.; Kim, K.; Kim, D.-H. Groundwater Level Prediction Model Using Correlation and Difference Mechanisms Based on Boreholes Data for Sustainable Hydraulic Resource Management. *IEEE Access* **2021**, *9*, 96092.

(19) Seidler, R.; Padalkina, K.; Bucker, H. M.; Ebigbo, A.; Herty, M.; Marquart, G.; Niederau, J. Optimal experimental design for finding optimally located boreholes in geothermal engineering. *Proceedings World Geothermal Congress* **2016**, *20*, 375.

(20) Rizwan, A.; Khan, A. N.; Iqbal, N.; Ahmad, R.; Kim, D. H. Enhanced Optimization-Based Voting Classifier and Chained Multi-Objective Regressor for Effective Groundwater Resource Management. *IEEE Access* **2021**, *9*, 168329−168341.

(21) Mantha, B.; Samuel, R. ROP optimization using artificial intelligence techniques with statistical regression coupling. *SPE annual technical conference and exhibition* **2016**, No. SPE-181382-MS.

(22) Baek, S.-S.; Pyo, J.; Chun, J. A. Prediction of Water Level and Water Quality Using a CNN-LSTM Combined Deep Learning Approach. *Water* **2020**, *12*, 3399. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.

(23) Al-AbdulJabbar, A.; Mahmoud, A. A.; Elkatatny, S. Artificial neural network model for real-time prediction of the rate of penetration while horizontally drilling natural gas-bearing sandstone formations. *Arabian Journal of Geosciences* **2021**, *14*, 1−14.

(24) Mosavi, A.; Sajedi Hosseini, F.; Choubin, B.; Goodarzi, M.; Dineva, A. A.; Rafiei Sardooi, E. Ensemble Boosting and Bagging Based Machine Learning Models for Groundwater Potential Prediction. *Water Resources Management* **2021**, *35*, 23−37.

(25) Su, J.; Zhao, Y.; He, T.; Luo, P. Prediction of drilling leakage locations based on optimized neural networks and the standard random forest method. *Oil & Gas Science and Technology − Revue d'IFP Energies nouvelles* **2021**, *76*, 24. Publisher: EDP Sciences.

(26) Burduk, R. Recognition Task with Feature Selection and Weighted Majority Voting Based on Interval-Valued Fuzzy Sets. *Computational Collective Intelligence. Technologies and Applications* **2012**, 204−209.

(27) Zukui, L.; Xinxu, Z.; Jinhai, Z.; Daixu, T.; Yongsong, Y.; Jing, Y.; Yanqing, L. The experiment investigation of the correlation of acoustic logging and rock mechanical and engineering characteristics. *ISRM International Symposium-2nd Asian Rock Mechanics Symposium* **2001**, No. ISRM-ARMS2-2001-016.

(28) Shokir, E. M. E. M.; Emera, M. K.; Eid, S. M.; Wally, A. W. A New Optimization Model for 3d Well Design. *Oil & Gas Science and Technology* **2004**, *59*, 255−266. Number: 3 Publisher: EDP Sciences.

(29) Gong, W.; Tien, Y.-M.; Juang, C. H.; Martin, J. R.; Luo, Z. Optimization of site investigation program for improved statistical characterization of geotechnical property based on random field theory. *Bulletin of Engineering Geology and the Environment* **2017**, *76*, 1021−1035.

(30) Rostamian, A.; Jamshidi, S.; Zirbes, E. The development of a novel multi-objective optimization framework for non-vertical well placement based on a modified non-dominated sorting genetic algorithm-II. *Computational Geosciences* **2019**, *23*, 1065−1085.

(31) Pereira, R. A.; Moura, A. V.; de Souza, C. C. Comparative experiments with GRASP and constraint programming for the oil well drilling problem. *International Workshop on Experimental and Efficient Algorithms* **2005**, *3503*, 328−340.

(32) Murugan, P.; Kannan, S.; Baskar, S. NSGA-II algorithm for multi-objective generation expansion planning problem. *Electric power systems research* **2009**, *79*, 622−628.

(33) Chang, D.-E.; Ha, K.-R.; Jun, H.-D.; Kim, J.-H.; Kang, K.-H. Determination of the Optimal Location for Water Treatment Plants in the Decentralized Water Supply System. *Journal of Korean Society of Water and Wastewater* **2013**, *27*, 1−10.

(34) Seidler, R.; Bücker, H.; Graf, K.; Herty, M.; Niederau, J.; Marquart, G.; Clauser, C. Optimal Experimental Design (OED) for Finding Optimally Located Boreholes in Geothermal Engineering. World Geothermal Congress, Melbourne, Australia, 19−25 April 2015.

(35) Daneshfar, M. A.; Ardjmand, M. A new approach in the optimal site selection of landfills for drilling cuttings from petroleum and gas fields. *Chemosphere* **2021**, *270*, 129402.
(36) Poursalehi, F.; Akbarpour, A.; Hashemi, S. R. Locating optimal position of artificial recharge wells in aquifer using grey wolf optimization algorithm and isogeometric numerical method. *Applied Water Science* **2022**, *12*, 164.

https://doi.org/10.1021/acsomega.2c06854
ACS Omega 2023, 8, 10806−10821