Ecological Indicators 127 (2021) 107790

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Evaluating the predictive power of different machine learning algorithms for groundwater salinity prediction of multi-layer coastal aquifers in the Mekong Delta, Vietnam

Dang An Tran [a,*], Maki Tsujimura [b], Nam Thang Ha [c], Van Tam Nguyen [d], Doan Van Binh [a,e], Thanh Duc Dang [f], Quang-Van Doan [g], Dieu Tien Bui [h], Trieu Anh Ngoc [a], Le Vo Phu [i,j], Pham Thi Bich Thuc [k], Tien Dat Pham [l]

[a] Faculty of Water Resources Engineering, Thuyloi University, 175 Tay Son, Dong Da, Hanoi, Viet Nam
[b] Faculty of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
[c] Faculty of Fisheries, University of Agriculture and Forestry, Hue University, 530000, Viet Nam
[d] Department of Hydrgeology, Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany
[e] Water Resources Center, Disaster Prevention Research Institute, Kyoto University, Goka-sho, Uji City, Kyoto 611-0011, Japan
[f] Institute for Water and Environment Research, Thuyloi University, Ho Chi Minh City, Viet Nam
[g] Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
[h] GIS group, Department of Business and IT, University of South-Eastern Norway
[i] Faculty of Environment and Natural Resources, Ho Chi Minh City University of Technology (HCMUT), 268, Ly Thuong Kiet street, District 10, HCMC, Vietnam
[j] Vietnam National University Ho Chi Minh City, Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Viet Nam
[k] Institute of Applied Mechanics and Informatics, Vietnam Academy of Science and Technology, 291 Dien Bien Phu Street, Ward 7, District 3, Ho Chi Minh City, Viet Nam
[l] Department of Biological Sciences, Florida International University (FIU), Miami 33199, USA

ARTICLE INFO

Keywords:
CatBoost Regression
Influencing factors
Groundwater salinization
Multi-layer coastal aquifers
Mekong Delta

ABSTRACT

Groundwater salinization is considered as a major environmental problem in worldwide coastal areas, influencing ecosystems and human health. However, an accurate prediction of salinity concentration in groundwater remains a challenge due to the complexity of groundwater salinization processes and its influencing factors. In this study, we evaluate *state-of-the-art* machine learning (ML) algorithms for predicting groundwater salinity and identify its influencing factors. We conducted a study for the coastal multi-layer aquifers of the Mekong River Delta (Vietnam), using a geodatabase of 216 groundwater samples and 14 conditioning factors. We compared the predictive performances of different ML techniques, i.e., the Random Forest Regression (RFR), the Extreme Gradient Boosting Regression (XGBR), the CatBoost Regression (CBR), and the Light Gradient Boosting Regression (LGBR) models. The model performance was assessed by using root-mean-square error (RMSE), coefficient of determination (R^2), the Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The results show that the CBR model has the highest performance with both training (R^2 = 0.999, RMSE = 29.90) and testing datasets (R^2 = 0.84, RMSE = 205.96, AIC = 720.60, and BIC = 751.04). Ten of the 14 influencing factors, including the distance to saline sources, the depth of screen well, the groundwater level, the vertical hydraulic conductivity, the operation time, the well density, the extraction capacity, the thickness of the aquitard, the distance to fault, and the horizontal hydraulic conductivity are the most important factors for groundwater salinity prediction. The results provide insights for policymakers in proposing remediation and management strategies for groundwater salinity issues in the context of excessive groundwater exploitation in coastal lowland regions. Since the human-induced influencing factors have significantly influenced groundwater salinization, urgent actions should be taken into consideration to ensure sustainable groundwater management in the coastal areas of the Mekong River Delta.

* Corresponding author.
E-mail addresses: antd@tlu.edu.vn (D.A. Tran), mktsuji@geoenv.tsukuba.ac.jp (M. Tsujimura), hanamthang@hueuni.edu.vn, hanamthang@huaf.edu.vn (N.T. Ha), tam.nguyen@ufz.de (V.T. Nguyen), binhdv@tlu.edu.vn (D.V. Binh), thanhiwer@gmail.com (T.D. Dang), doan.van.gb@u.tsukuba.ac.jp (Q.-V. Doan), dieu.t.bui@usn.no (D.T. Bui), ngocta@tlu.edu.vn (T. Anh Ngoc), volephu@hcmut.edu.vn (L.V. Phu), ptbthuc@iami.vast.vn (P.T.B. Thuc), dat6784@gmail.com (T.D. Pham).

https://doi.org/10.1016/j.ecolind.2021.107790
Received 5 July 2020; Received in revised form 15 January 2021; Accepted 5 May 2021
Available online 17 May 2021

1. Introduction

As high yield, good quality, easy extraction, and operation at acceptable prices, groundwater has become a primary freshwater resource for approximately two billion people, and it accounts for around one-third of the total water withdrawal worldwide (Aeschbach-Hertig and Gleeson, 2012; Famiglietti, 2014). However, many coastal areas across the world are experiencing freshwater scarcity due to seawater intrusion into the surface water system. Therefore, groundwater in these regions has become an alternative freshwater source for drinking, industries and agriculture (Behera et al., 2019; Kagabu et al., 2011; Kaur et al., 2020; Mohanty and Rao, 2019; Singh et al., 2018). Despite its essential role, groundwater resource is highly vulnerable to overexploitation, seawater intrusion, climate change, and sea-level rise (Akbari et al., 2020; Ferguson and Gleeson, 2012; Singh, 2015). Saline groundwater in coastal aquifers may originate from paleo-seawater intrusion (Delsman et al., 2014; Gossel et al., 2010; Larsen et al., 2017), modern seawater intrusion (Yechieli et al., 2019), leaking brines from oil fields (Chekirbane et al., 2014), and irrigation return flows (Foster et al., 2018). High salinity concentration in water could have severe impacts on ecosystems (Korres et al., 2019; Ma et al., 2019; Radanielson et al., 2018; Velasco et al., 2019) and could cause many health-related problems (Akter, 2019; Nahian et al., 2018; Park and Kwock, 2015; Rahaman et al., 2020; Shammi et al., 2019; Vineis et al., 2011). For example, exceeding salinity in irrigated water may reduce crop yields (Korres et al., 2019; Parvin et al., 2019; Radanielson et al., 2018); a high intake of salinity causes more stress on the development of organisms in fresh ecosystems (Velasco et al., 2019). Meanwhile, using drinking water with high salinity leads to an increase of hypertension (Shammi et al., 2019; Vineis et al., 2011), coronary heart disease (Park and Kwock, 2015; Rahaman et al., 2020), and chronic kidney disease (Naser et al., 2017; Rahaman et al., 2020). Thus, insights into salinity intrusion mechanisms and providing a reliable prediction of groundwater salinity are crucial for providing an adequate water source for diverse sectors and human health protection (Carretero et al., 2013; Gejl et al., 2020; Kaur et al., 2020; Larsen et al., 2017).

There are several methods to investigate salinization processes in coastal aquifers, e.g., physical experiments, physically-based models, and data-driven prediction models (Boluda-Botella et al., 2008; Crestani et al., 2019; Guo et al., 2019; Rajaee et al., 2019; Robinson et al., 2018; Yadav et al., 2018; Yousefi et al., 2020; Yu et al., 2019). Physical experiments provide an understanding of saltwater movement processes by measuring solute concentration in a specific period (Crestani et al., 2019). This method is often combined with physical-based models to gain a better understanding of saltwater movement (Boluda-Botella et al., 2008; Guo et al., 2019; Yu et al., 2019). Although physical experiments are useful to understand seawater intrusion processes in a short period, these techniques are only applied to investigate a simple process of seawater intrusion while the processes of seawater intrusion, in reality, is very complex (Badaruddin et al., 2017; Yu and Michael, 2019). Physically-based models (e.g., MODFLOW, MT3DMS, SEAWAT, FEMWATER) are alternatives to examine seawater intrusion into coastal multi-aquifers systems in the context of human activities and climate change. In recent decades, physically-based models have been widely used for modelling groundwater dynamics and seawater intrusion into coastal aquifers (Abdelhamid et al., 2016; Lal and Datta, 2019; Mahmoodzadeh and Karamouz, 2019; Stein et al., 2019; Voss and Souza, 1987; Yu and Michael, 2019). These models are based on a detailed process description (e.g., Richards and convection–dispersion equations) and can provide a mechanistic understanding of physical processes not only at a small scale but also at a regional scale (Larsen et al., 2017; Pham and Tsai, 2016). The applications of such models, however, are not always possible because they often require extensive input data and computational demand. Therefore, many researchers have developed and applied machine learning methods for groundwater studies because these methods require fewer input data and computational time while providing comparable results (Sajedi-Hosseini et al., 2018).

In recent years, machine learning methods have been proven as promising tools for estimating groundwater levels (Chen et al., 2020; Rahman et al., 2020a; Sahoo et al., 2017), mapping potential groundwater recharge (Pourghasemi et al., 2020), assessing groundwater vulnerability (Moazamnia et al., 2020; Moghaddam et al., 2020) and predicting groundwater quality (Li et al., 2020; Liu et al., 2017; Rahmati et al., 2019). For instance, Kopsiaftis et al. (2019) applied the Gaussian process regression (GPR), the regression trees (RT), and the Support Vector Machine regression (SVR) models to predict the extension of seawater intrusion in a coastal area. The result showed that the GPR model had a better predictive power compared to those obtained by the RT and the SVR techniques. Yadav et al. (2018) tested different data-based models, including the artificial neural network (ANN), the SVR, the genetic programming (GP), and the extreme learning machine (ELM) algorithms to develop three-dimensional, density-dependent flow and transport processes in a coastal aquifer system. It was concluded that the SVR model provided the most accurate prediction compared to other models. Lal and Datta (2019) illustrated the capacity of the GPR model in estimating density-dependent saltwater intrusion processes and predicting salinity concentrations in an exemplary coastal aquifer system. Robinson et al. (2018) applied the random forests regression (RFR) algorithm to optimize calibration using image-based mapping seawater intrusion. Recently, the ensemble-based learning techniques such as the extreme gradient boosting regression (XGBR), the CatBoost Regression (CBR), and the Light Gradient Boosting Regression (LGBR) algorithms have shown the advancement in dealing with non-linear processes, providing the high potentiality in predicting natural phenomena. For instance, the XGBR model has illustrated as the highly accurate prediction model for environmental indicators such as streamflow (Ni et al., 2020; Yu et al., 2020), forest above-ground biomass estimation (Pham et al., 2020a, 2020b), solar radiation (Fan et al., 2018), and air pollution (Chen et al., 2019; Pan et al., 2019). More recently, the CBR model developed by the Yandex Company (Liudmila Prokhorenkova et al., 2017) has demonstrated as a potential tool for accurately predicting different environmental indicators (Dev and Eden, 2019; Huang et al., 2019; Matsuzaka et al., 2020; Zhang et al., 2020). Likewise, the LGBR model has performed well in predicting different environmental indicators, such as evapotranspiration and temperature (Fan et al., 2019; Gong et al., 2020). The abovementioned studies demonstrated that machine learning techniques are highly flexible ability to handle complex non-linear environmental problems; however, the model performance of different ML models varies widely, depending on selected algorithms, influencing factors, and targeted predictors. Therefore, it is necessary to evaluate different algorithms in combination with various influencing factors to choose an appropriate machine learning model for specific environmental indicators.

Groundwater salinization in coastal areas, which reflects both past and present seawater intrusion processes in coastal aquifers, is generally caused by the direct movement of seawater into fresh-aquifers (Han and Currell, 2018; Larsen et al., 2017). Besides, saline groundwater in coastal aquifers shows a spatial heterogeneity, depending on geological features, aquifer properties, hydrological processes (Shi et al., 2020), and especially human activities (Scharping et al., 2018; Walther et al., 2020). This leads to a significant challenge to an accurate prediction of groundwater salinity and the identification of its influencing factors. This study aims to (1) evaluate the potential application of the novel ensemble-based machine learning techniques (XGBR, CBR, and LGBR) to predict groundwater salinity in comparison with the most widely used algorithms (RFR), (2) to improve model prediction using feature selection functions, and (3) to investigate whether the machine learning approaches could provide a physically-sound understanding of salinity intrusion mechanisms in coastal aquifers. To the best of our knowledge, the current study is the first time in which the RF, the XGB, the CBR, and the LGBR machine learning algorithms were evaluated to select the most suitable data-driven model for accurately predicting groundwater

salinization and identifying its influencing factors in a coastal multi-aquifer system. In this study, the Mekong Delta (MD), Vietnam, which was taken as a representative case study, is one of the most vulnerable coastal areas exposed to integrated impacts of human activities and climatic variation (Binh et al., 2020; Shrestha et al., 2016; Tran et al., 2020).

2. Study area and data preparation

2.1. Description of the study area

The Mekong Delta (MD) is located in the Lower Mekong Basin, which is one of the most important deltas for national and regional food security and biodiversity conservation (Ha et al., 2018). However, this delta has been facing severe water scarcity due to seawater intrusion and water pollution, especially in the coastal region where most of the freshwater demand depending on groundwater sources (Hamer et al., 2020; Tran et al., 2020). The study area is located in the southeastern region of the Mekong Delta (MD), Vietnam (Fig. 1). It covers an area of 3,312 km^2 with the elevation ranging from 0.5 to 2.5 m above the mean sea level.

The study area borders the Hau River to the Northwest and the East Vietnam Sea to the Southwest. This study site has approximately 1.20 million people, most of whom depend on agriculture for their livelihoods, contributing to 42% of the total GDP of the province. Agriculture and aquaculture lands are the dominant land use types, accounting for 84.8% (276,690 ha of the total area), which includes rice fields (52.98%), aquaculture (19.7%), orchards (15.5%), lands of other vegetable types (6.8%), and other types of land use (5.1%) (Data source: Decision No.108/NQ-CP of the Government 2018). Groundwater in the area is a dominant source of water for domestic, industrial, and agricultural activities, and the long-term extraction of groundwater has resulted in a groundwater level depletion in irrigated regions (Minderhoud et al., 2018). Groundwater salinization has been identified as one of the most significant threats to the sustainable development of the region (Tran et al., 2020). The extent of groundwater salinization in the

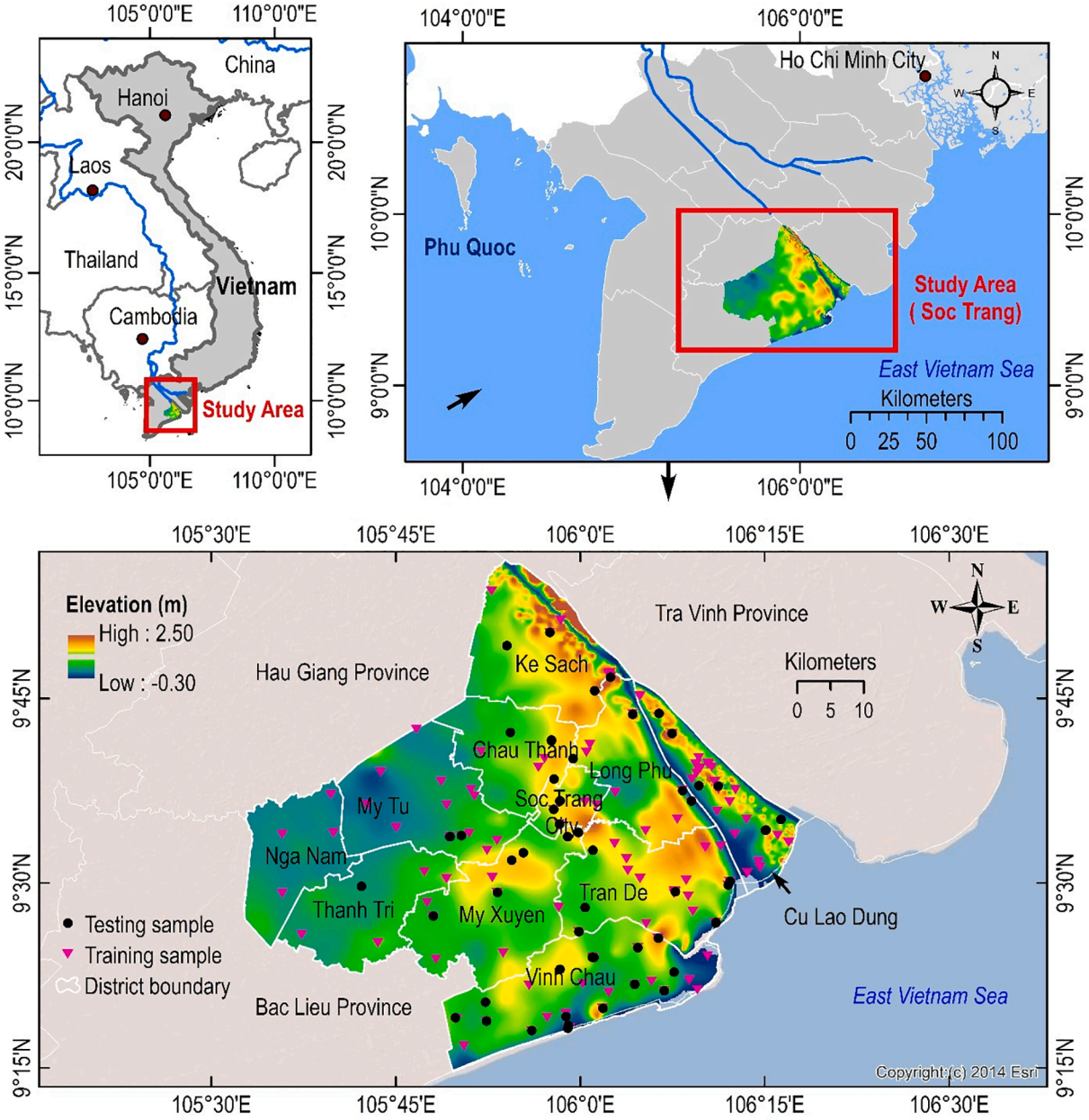

Fig. 1. Location of the study area and sampling wells.

study area has recently increased due to the rapid increase in groundwater demand (Minderhoud et al., 2017; Nam et al., 2019). The study area has a dense river system with a direct connection to the sea. The hydrological regime of the study area is complex and strongly influenced by the flow regime of the Mekong river and tidal fluctuation (Binh et al., 2020). The study area is in a tropical monsoon climate region with two distinct seasons, the dry season (May to November) and the rainy season (December to April). The annual average rainfall is about 1,772 mm with substantial seasonal variations, in which approximately 85% of the annual rainfall occurs during the rainy season. The study area has also been recognized as one of the most vulnerable regions to climate change and sea-level rise in the world (Dang et al., 2018; Shrestha et al., 2016; Smajgl et al., 2015).

The hydrogeological setting of the study area is characterized by a multi-layered aquifer system, formed between the Miocene and Holocene epoch (Hung Van et al., 2019; Wagner et al., 2012). Groundwater in the Pleistocene aquifers is the primary source of drinking water because these aquifers have high yield and good-quality water compared to other aquifers (An et al., 2018). In this study, we aim at predicting salinity concentration in the middle and lower Pleistocene aquifers and identifying the main influencing factors of this process.

2.2. Data preparation

A total of 216 groundwater samples was collected from the Pleistocene aquifers during both the rainy and the dry seasons between 2013 and 2018. On-site measurements were conducted to obtain physical parameters such as groundwater temperature (T °C), pH, dissolved oxygen (DO), and electrical conductivity (EC), using the HANNA portable instruments (Hanna Instruments Inc. 2015). Chloride concentration in groundwater samples was analyzed by using Ion Liquid Chromatography (Shimadzu Co. Ltd., Japan) at the University of Tsukuba, Japan. Fourteen influencing factors were prepared (Table 1) based on a geological map, a drainage network, and *in-situ* data from the study area (Fig. 2).

Table 1
Geospatial data sources used for the salinity prediction in this research.

No.	Explanatory Variables	Abbreviation	Unit	Sources
1	Distance to saline sources	DTS	km	DONRE
2	Depth of screen well	DSW	km	DONRE
3	Groundwater level	GWL	m.a.m.s.l	This study
4	Vertical hydraulic conductivity	VHC	m/d	DWRPIS 2010
5	Operation time of well	OOW	year	DONRE
6	Well density	WED	well/km^2	DONRE
7	Extraction capacity	EXC	m^3/d	DONRE
8	Thickness of aquitard	TOA	m	DONRE
9	Distance to fault	DTF	km	DONRE
10	Horizontal hydraulic conductivity	HHC	m/d	Pumping test, DWRPIS 2010
11	Distance to the sea	DTS	km	Drainage network, Vietnam Map 2008
12	Distance to hydraulic window	DHW	km	Drainage network, Vietnam Map 2008
13	Extraction density	EXD	m^3/km^3	Drainage network, Vietnam Map 2008
14	Soil type	ST		DONRE

DONRE: Department of Natural Resources and Environment, Soc Trang Province, Vietnam.
DWRPIS: Division of Water Resources Planning and Investigation in South of Vietnam.

3. Methodology

3.1. Machine learning models

3.1.1. Random forest (RF)

RF is a well-known ML method, which has been widely applied for both classification and regression problems based on the ensemble of decision trees (Breiman, 2001). A decision tree is a top-down tree-like structure, in which each non-leaf node is a test, each branch is an outcome of the test, and each leaf node is a decision. Regression with a single decision tree may result in overfitting (high variance) and depends on the distribution of training sets. A large number of decorrelated decision trees can form a random forest, which then can reduce the variance and boost model performance (Criminisi, 2011). The procedure forming RF is as follows: (1) *n* random subsets (called "bootstrapped subsets") are sampled from a training dataset, and this process is based on a random selection of features of the dataset. A subset may contain overlapped data in other subsets; (2) *n* decision trees are built using these *n* bootstrapped subsets (Fig. 1). The number of trees *n* is decided by using either cross-validation or out-of-bag (OOB) error methods. A detailed description of the statistical formulation of the RF algorithm can be found in Breiman (2001).

3.1.2. Extreme gradient boosting (XGB)

XGB is an ensemble-machine learning algorithm that is based on decision trees (Friedman, 2001). However, a boosting model constructs the "forest" of decision trees sequentially, or one decision tree can be built on learning experience inherited from previous trees (Chen and Guestrin, 2016; Johnson et al., 2018). The second tree focuses on the cases in which the first tree gives a poor prediction, and this learning process is repeated much time in such a way that the combination of these trees can better capture the relationship between predictands and predictors. Gradient boosting is a form of boosting models in which weak prediction cases are assessed if they contribute to minimize the overall lost function (also called the prediction error) (Lim and Chi, 2019). An ensemble can be considered as highly valuable if the added decision tree built for this case can significantly reduce the prediction error while no change in the error implicates a no value case; thus, only useful decision trees are kept. This may give the XGB model advantages in handling complex problems such as quantifying saline concentration in groundwater since data measurement in the underground environment may contain many exceptional cases. Note that model parameters control the learning efficiency of each machine learning algorithm, and in the case of the XGB model, they include three groups: tree-specific, boosting, and different metrics. The selection of these model parameters is a challenging task and depends on user experience while this process does not always return in an optimum set of parameters. Thus, we employed a grid-search algorithm with five-fold cross-validation for hyperparameter tuning to improve the accuracy of the XGB model.

3.1.3. CatBoost (CB)

CB is a novel ensemble-based learning algorithm and can be applied for either the categorical or numerical data types in the multi-task of regression, ranking, binary and multiclass classification (Liudmila Prokhorenkova et al., 2017). The name of CatBoost was combined from "category" and "boosting", giving a description of the algorithm in processing category data. The CB algorithm, which uses an efficient preprocessing of categorized data, is Target-Based with prior statistics (TBS) and helps to reduce target leakage. The CB model implements ordered boosting, a permutation driven boosting method, to avoid overfitting (by adjusting a low learning rate) for small datasets. This ML technique uses the oblivious decision tree with the same splitting criteria applied across the entire tree, and this decision tree is different from the asymmetric tree in the XGB or LGB algorithms. The algorithm is remarkably faster compared to other models in the training speed because it can be supported by Graphical Processing Unit (GPU), and it is robust and requires fewer hyperparameters to be tuned.

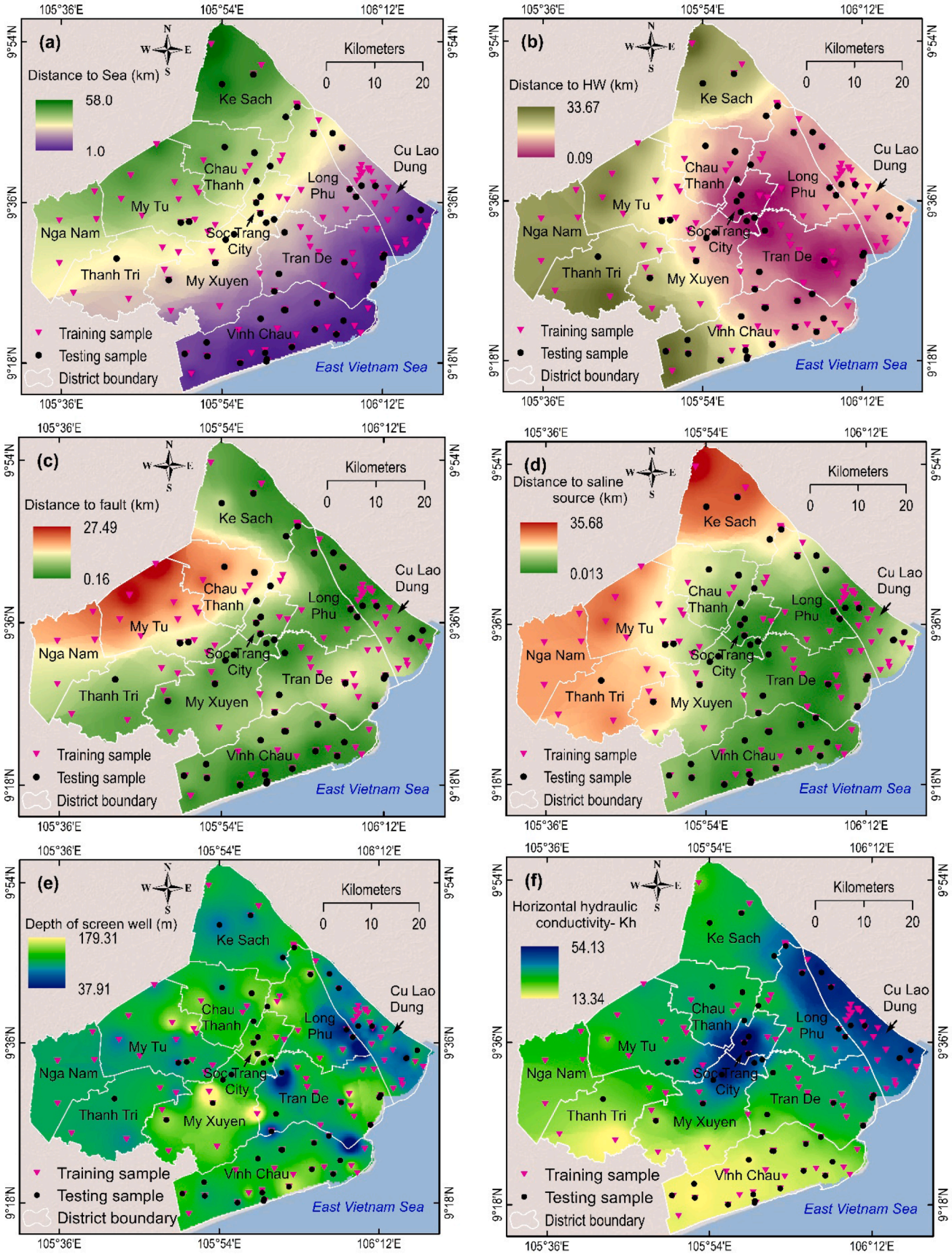

Fig. 2. Influencing factors of groundwater salinization: (a) distance to sea; (b) distance to hydraulic window; (c) distance to fault; (d) distance to saline sources; (e) depth of screen well; (f) horizontal hydraulic conductivity (K_h), (g) vertical hydraulic conductivity (K_v); (h) thickness of aquitard; (i) operation time; (j) well density; (k) extraction density; (l) extraction capacity; (m) groundwater level; and (n) soil type.

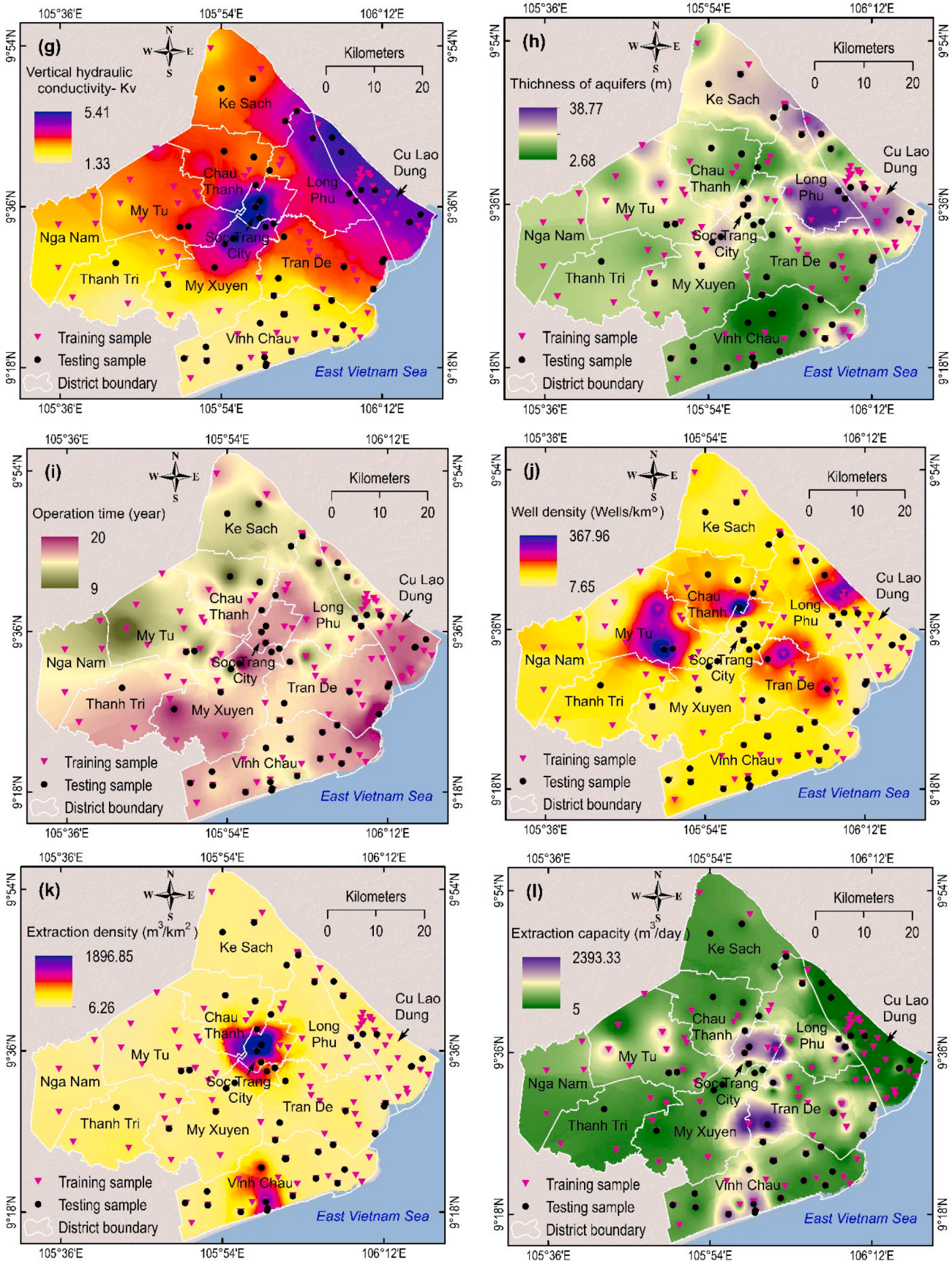

Fig. 2. (*continued*).

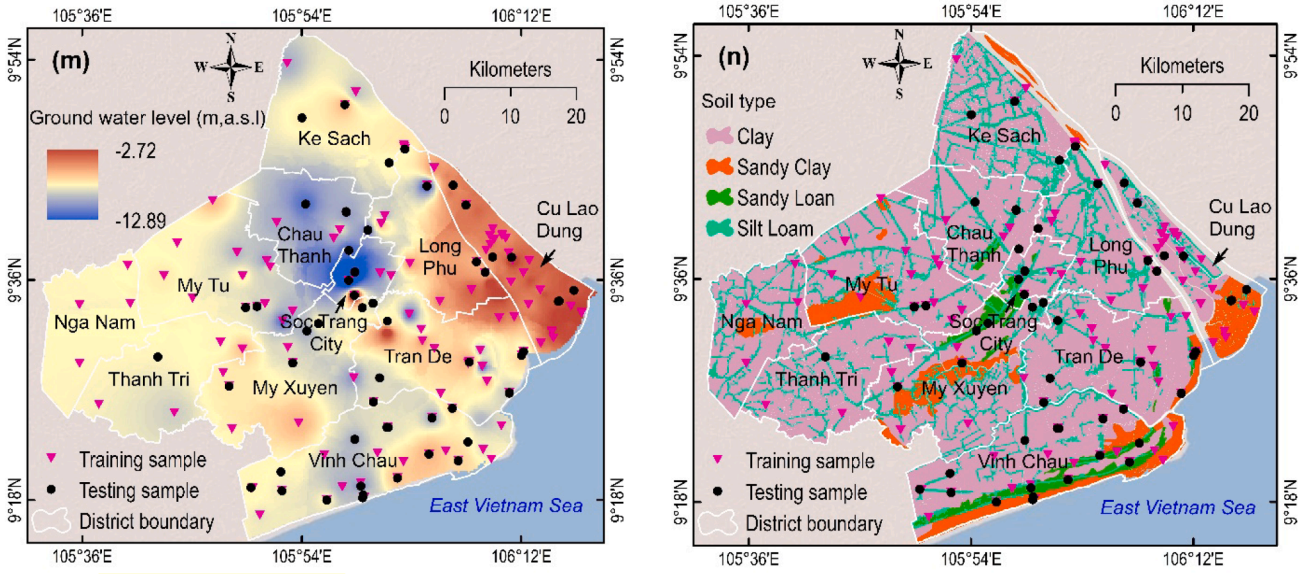

Fig. 2. (*continued*).

3.1.4. Light gradient boosting (LGB)

The LGB algorithm is an innovative machine learning (ML) that was initially released in 2017 (Ke et al., 2017). This algorithm is designed for a variety of tasks, including regression, binary classification, multiclass classification, and lambda classification. This technique is a decision tree-based algorithm with the gradient boosting machine inside. It is, however, different from other ML decision tree-based algorithms in growing the tree vertically (leaf-wise). The LGB algorithm selects the leaf with a maximum delta loss rather than horizontally (level-wise). To improve the model accuracy, LGB uses two novel gradient boosting techniques, i.e., Gradient-Based One-Side Sampling (GOSS) and Dropouts meet Multiple Additive Regression Tree (DART), which can handle complexity in computation and large datasets. The LGB model is proposed to be highly efficient for breakneck training speed and low memory demand. The algorithm, however, comes with a hundred of parameters that need to be tuned and maybe overfitted with small datasets.

3.2. General approach

Our first goal was to investigate the applicability of machine learning models in predicting groundwater salinization concentration in the MRD, Vietnam. For this purpose, we evaluated and compared the four machine learning models. The model was then improved with the optimal conditioning factors by using a feature selection process. The overall computational framework is illustrated in Fig. 3.

3.2.1. Data preprocessing

The accumulation of salinity in groundwater is a complex process because of different influencing factors (Kanagaraj et al., 2018; Mahlknecht et al., 2017). Thus, the influencing factors for groundwater salinization prediction are selected by analyzing characteristics of seawater intrusion processes into fresh coastal aquifers (Han and Currell, 2018; Tran et al., 2019). In the Pleistocene aquifers, groundwater salinity originates from (1) downward or upward leakage of paleo-saline water (Chatton et al., 2016; Khaska et al., 2013), (2) halite dissolution in the topsoil layers (Blasco et al., 2019; Walter et al., 2017), (3) seawater intrusion (Han and Currell, 2018; Kanagaraj et al., 2018), and (4) irrigation return flow (Essaid and Caldwell, 2017; Lapworth et al., 2017; Tweed et al., 2018). The downward or upward leakages of paleo-saline water, which are related to aquifer properties, are further incorporated into the soil properties influencing factors. Furthermore, the thicknesses

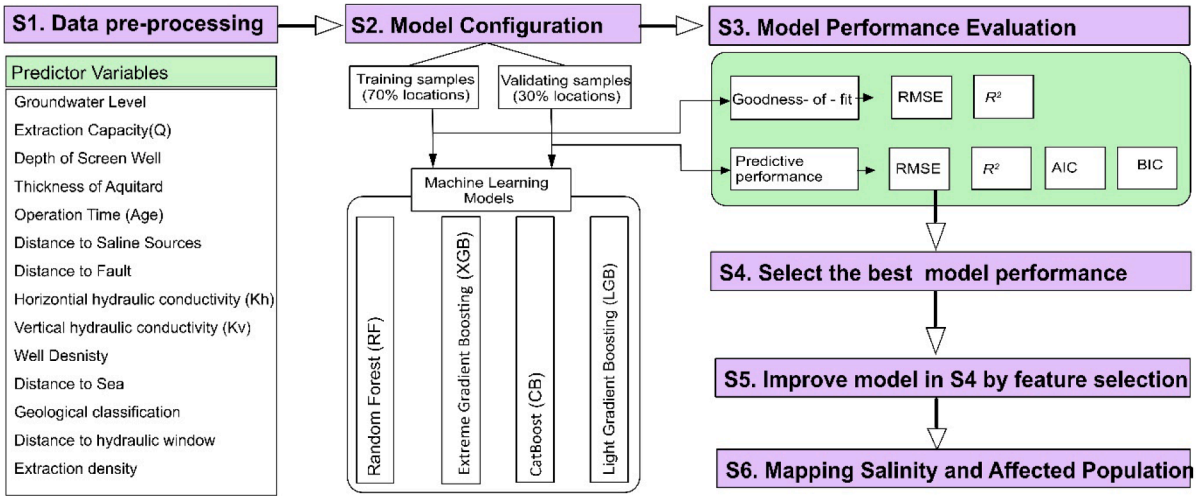

Fig. 3. The proposed modelling framework used in this study.

of aquitards, the distance to hydraulic windows, the distance to faults, and the hydraulic conductivity could also affect the leaking rate (Elmahdy and Mohamed, 2013). Also, other geographical variables such as the distance from main rivers, the distance to drainage, and the drainage density are widely considered as the influencing factors to groundwater salinity (Winkel et al., 2008). Halite dissolution depends on salt rock/sediment properties and horizontal hydraulic conductivity. Variables which represent the impacts of human activities on groundwater salinity in the study area consist of the groundwater level, the extraction capacity, the well density, the extraction density, and the operation time. The severity of seawater intrusion may also depend on the distance to saline sources to faults (Yechieli et al., 2019). The interaction between these four processes results in a complex salinization process in the study area (Tran et al., 2020). In this study, the 216 groundwater samples (each sample consists of fourteen variables) from the middle and lower Pleistocene aquifers were used (Table 1). The measured Cl^- concentration is assigned as a dependent variable, and the fourteen influencing factors are set as independent variables.

3.2.2. Model configuration and evaluation

3.2.2.1. Machine learning (ML) model hyper-parameter tuning. The ML models consist of various hyper-parameters that should be tuned before the implementation to attain the best model performance. In this study, we tuned the hyper-parameters of selected ML models using a grid search function with five-fold cross-validation in Scikit-learn packages (Pedregosa et al., 2012). During the grid search processing, several combinations of the model's hyper-parameters were considered and tested until reaching a minimum RMSE score. The grid search returned the best combination of hyper-parameters for each ML model.

3.2.2.2. Model configuration and training. The configuration and the training phases of the four ML models are conducted in Python version 3.7. The 216 groundwater samples were randomly divided into a training dataset (70%) and a testing dataset (30%) to evaluate model performance by using the Scikit-learn packages (Pham et al., 2020a). The hyperparameter of the four ML models (RFR, XGBR, CBR, and LGBR) was tuned by using a grid search function with five-fold cross-validation in Scikit-learn packages (Table 2). Then, the model with the highest predictive performance (i.e., the highest R^2 and the lowest RMSE, AIC, and BIC values) in both the training and the testing phases was selected. Once the best model is determined, we employed the selected model to predict salinity concentration (Chloride in mg/L) in each well location.

3.2.2.3. Performance assessment. The performance criteria used for evaluating model performance depend on the output variables of each model, e.g., categorical or continuous variable (Tien Bui et al., 2016). Performance criteria such as the root-mean-square error (RMSE, Eq. (1)) and the coefficient of determination (R^2, Eq. (2)), were used in this study for evaluating the models with continuous output values. Moreover, the Taylor diagram was used to compare the predictive performance of the models (Rahman et al., 2020b; Taylor, 2001). Each performance criterion indicates specific information regarding predictive performance efficiency (Li et al., 2016; Tien Bui et al., 2018).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i^e - y_i^m)^2}{N}} \tag{1}$$

$$R^2 = \frac{\sum_{i=1}^{N}(y_i^e - \overline{y^e}) \times (y_i^m - \overline{y^m}))}{\sqrt{\sum_{i=1}^{N}(y_i^e - \overline{y^e})^2 \times (y_i^m - \overline{y^m})^2}} \tag{2}$$

where y_i^e and y_i^m are predicted and measured chloride (Cl^-) concentration in observation i, respectively, and N is the number of observations, and $\overline{y^e}$and $\overline{y^m}$ are the mean values of predicted Chloride (Cl^-) and measured Chloride (Cl^-), respectively. Higher values of R^2 are preferred, i.e., closing to 1 means better model performance, and the regression line fits the training/testing data well. Conversely, the lower values of RMSE indicates better model performances.

Also, the Akaike Information Criterion (AIC) (Akaike, 1974) and the Bayesian Information Criterion (BIC) (Stone, 1979) were used in this study to select the best model for predicting groundwater salinity. AIC is a technique based on the sample fit to estimate the likelihood of a model to predict or estimate values (Burnham and Anderson, 2004). Meanwhile, the BIC measures the trade-off between model fit and complexity of the model. The lowest AIC and BIC values among models reveal the best model (Pham et al., 2017; Vrieze, 2012). The AIC and BIC values of a model are estimated by using the following Equations:

$$AIC = 2k - 2\log(L) \tag{3}$$

$$BIC = 2\log(n)k - 2\ln(L) \tag{4}$$

where k is the number of independent variables ($k = 14$, Fig. 2), L is the value of the likelihood, and n is the number of recorded measurements.

Since the influencing factors for predicting groundwater salinization have significantly different ranges, normalization was used to convert the values of numeric columns into a scale from 0 to 1 by using the following equation (Tien Bui et al., 2012; Wang and Huang, 2009):

$$x_n = \frac{x_o - x_{min}}{x_{max} - x_{min}} \tag{5}$$

where x_n and x_o represent the normalized and original data, respectively,x_{max} and x_{min} are the max and min original values, respectively.

3.2.2.4. Feature importance. The feature importance (variable importance) in ensemble-based decision tree algorithms is calculated by using the variable-importance approach (Dorogush et al., 2018; Prokhorenkova et al., 2018). In this approach, the model first searches for a candidate subset of variables from all variables with the grid search approach. The CBR model ranks the variables in descending order of their importances based on the root-mean-square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2). Then, a certain number of the least important variables are removed and the remaining variables form a variable subset. In this paper, the search/selection iterations were terminated when R^2 of the prediction model of the subset did not improve the performance with the testing dataset. In the final step, the selected variable subset was validated (in this case, by the 5-fold CV approach). Catboost Algorithms calculates variable importance using the below equation (Kang et al., 2019):

Table 2
Hyper-parameters of selected models.

Random Forest		**Extreme Gradient Boost**	
Bootstrap	True	Booster	Gbtree
Max depth	10	Gamma	5
Max features	3	Learning rate	0.01
Min sample leaf	4	Max depth	10
Min sample split	9	Min child weight	1
Number of trees	50	Number of trees	100
LightGBM		**CatBoost**	
Boosting type	Dart	Depth	7
Learning rate	0.1	Learning rate	0.209
Max depth	−1	Number of trees	100
Number of leaves	10		
Number of trees	50		

$$FI = \sum_{tree,leafsF}^{n} (v_1 - avr)^2 . c_1 + (v_2 - avr)^2 . c_2$$
$$avr = \frac{v_1 . c_1 + v_2 . c_2}{c_1 + c_2} \quad (6)$$

where FI is feature importance; c_1 and c_2 are the total weights of objects in the left and right leaves, respectively. This weight is equal to the number of objects in each leaf if weights are not specified for the dataset; and v_1 and v_2 represent the formula values in the left and right leaves, respectively. Calculating the modeling and generated variable importance of the CBR model was implemented in Python by using the library available at https://catboost.ai/docs/.

3.2.3. Mapping groundwater salinity and salinity-affected population

Firstly, we predicted chloride concentration and created a chloride map with a grid size of 300 × 300 m using the best-trained machine learning model (CB model) with the selected hyper-parameters and the top ten important variables. Secondly, the predictive chloride map was reclassified by using the drinking water standard from the World Health Organization (WHO) with five classes, including very low ($Cl^- < 250$ mg/L), low ($250 < Cl^- < 500$ mg/L), moderate ($500 < Cl^- < 1{,}000$ mg/L), high ($1{,}000 < Cl^- < 2{,}500$ mg/L), and extreme high (>2,500 mg/L). Thirdly, the salinity-affected area for each class of salinity concentration in groundwater was calculated by using the shoelace formula in ArcGIS 10.3 (Braden, 1986; Langenheim et al., 2017). Finally, the numbers of people within each area affected by salinity were estimated from the salinity-affected areas and population density.

4. Results and discussion

4.1. Model performance evaluation and comparison

The results in Table 3 show the goodness-of-fit and predictive performance of the four ML models in terms of the RMSE and R^2 metrics. In this study, the predictive models for groundwater salinization were constructed by using the training dataset and validated by using the testing dataset, drawing upon a total of 216 observation wells and the 14 variables. The goodness-of-fit indices show that the CBR model had the highest predictive performance, followed by the XGBR model. Conversely, the RFR and the LGBR models had the lowest performance to predict the chloride variations in the study site.

In the training phase, the CBR model has a better prediction of groundwater-salinity concentrations (RMSE = 29.90, R^2 = 0.99) than those of the XGBR model (RMSE = 336.90, R^2 = 0.66), the LGBMR model (RMSE = 485.95, R^2 = 0.29), and the RFR model (RMSE = 445.35, R^2 = 0.40). Although the goodness-of-fit performance reveals how well the model fits the training samples, it does not present the prediction and generalization capability of models (Rahmati et al., 2019). The predictive performance in the testing phase reflects how well the model can provide an accurate prediction (Table 3).

The results show that the CBR model had the highest predictive performance with the testing set (RMSE = 205.96, R^2 = 0.84), and the metrics were significantly better than those of the XGBR model (RMSE = 323.18, R^2 = 0.62), the LGBR model (RMSE = 472.89, R^2 = 0.19), and the RFR model (RMSE = 442.91, R^2 = 0.29). The visualization of the ML models' performance done by using the Taylor diagram also confirmed these results (Fig. 4).

The predicted values from the CBR and the XGBR models had higher correlations and lower RMSE with the observed chloride concentrations as compared to the LGBR and the RFR models. Besides, the AIC and the BIC values indicated a statistically significant difference between the four ML models (Table 3). Although we have considered various influencing factors to provide the accurate prediction of groundwater salinity in a coastal area of the MRD, the processes of seawater intrusion into fresh aquifers depend not only on natural variations but also on human activities.

Table 3
Model performance for chloride prediction.

Model	Training		Testing			
	R^2	RMSE	R^2	RMSE	AIC	BIC
RFR	0.40	445.35	0.29	442.91	820.13	850.58
XGBR	0.66	336.90	0.62	323.18	779.16	809.60
CBR	0.99	29.90	0.84	205.96	720.60	751.04
LGBMR	0.29	485.95	0.19	472.89	828.65	859.09

4.2. Feature selection

In the study area, among the original fourteen factors, the ten most important factors were selected. They (with the variable importance index) are the distance from saline sources (20.67), the deep of screen well (15.61), the groundwater level (14.02), the vertical hydraulic conductivity (10.30), the operation time of well (6.68), the well density (5.59), the extraction capacity (5.04), the thickness of aquitards (4.88), the distance to fault (4.38), and the horizontal hydraulic conductivity (4.14) (Fig. 5). The results reveal that groundwater salinization depends not only on the hydrogeological features (distance to saline sources, depth of screen well, vertical and horizontal hydraulic conductivities, and thickness of aquitard) but also groundwater extraction practices (groundwater level, operation time of well, extraction capacity, and well density).

These influencing factors also play an essential role in the transportation processes of other solutes such as arsenic, fluoride, and nitrate in groundwater (Podgorski et al., 2018; Ransom et al., 2017; Winkel et al., 2008). The hydrogeological features influence the leakage rate of saline groundwater from shallow to deeper aquifers (Hung Van et al., 2019) while groundwater exploitation activities exacerbate groundwater salinization (An et al., 2018). The result may also suggest that saline groundwater leaking from upper layers to lower layers is a dominant process, resulting in an increase of chloride concentration in groundwater of the study area. Hydraulically, an increased hydraulic gradient due to groundwater depletion coupled with high vertical hydraulic conductivity, the thin and discontinuous aquitard layer, and high-density gradients cause an increase of vertical flow rates (Ma et al., 2015). The similar findings which were also observed in other coastal aquifers in the world (Cary et al., 2015; Chatton et al., 2016; Delsman et al., 2014) indicated the strong influences of groundwater over-exploitation on seawater intrusion in coastal aquifers (Han et al., 2015; Larsen et al., 2017; Yechieli et al., 2019; Yu and Michael, 2019). The other important influencing factors have variable importance values from 3.0 (distance to the sea) to 0.42 (soil properties). It is noted that these indicators have lower values (<4.0), indicating less contribution to groundwater salinization processes. Meanwhile, the CBR model with the ten selected influencing factors (variable importance value > 4.0) produced a better prediction with RMSE = 193.51 and R^2 = 0.86. This result may suggest that paleo-saline groundwater salinization is the main process of increasing salinity concentration in this study while direct seawater intrusion from the sea to coastal aquifers is not dominant.

4.3. Mapping groundwater salinity

Since the CBR model has the most accurate performance in this study, we selected this model to map salinity concentration in the coastal multi-aquifers of the MRD. The result shows that the main salinity-affected region extends from the seaside to the central areas (Fig. 6).

The prediction results are consistent with the salinity observations in the study area reported in Tran et al. (2020). Accordingly, the high chloride concentrations which exceed the limit in the WHO drinking water standard ($Cl^- > 250$ mg/L) are predicted in the areas with paleo-saline sources, high extraction rates, and significant groundwater level

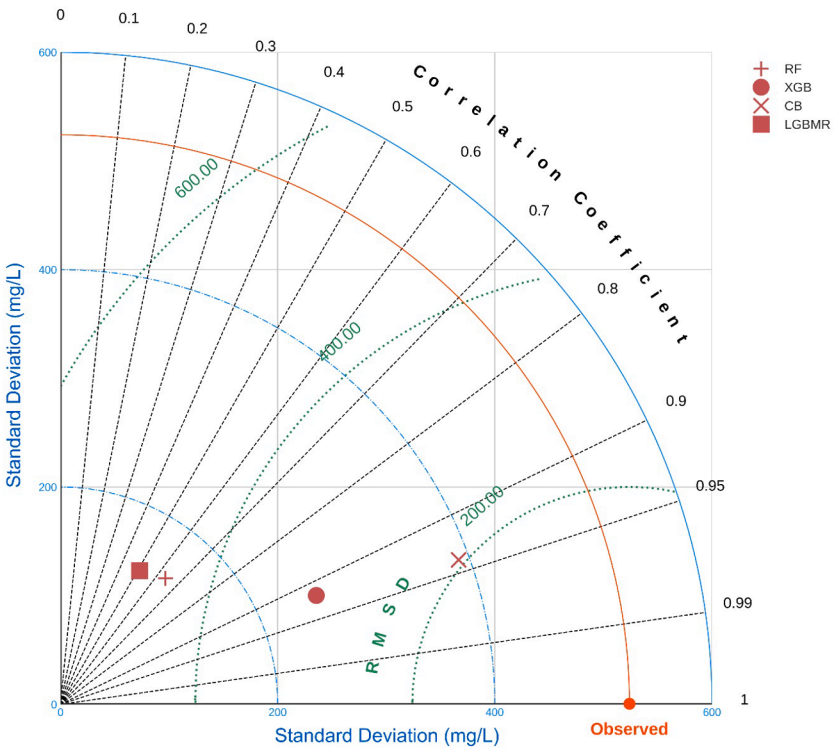

Fig. 4. Comparison of model's performances using the Taylor diagram.

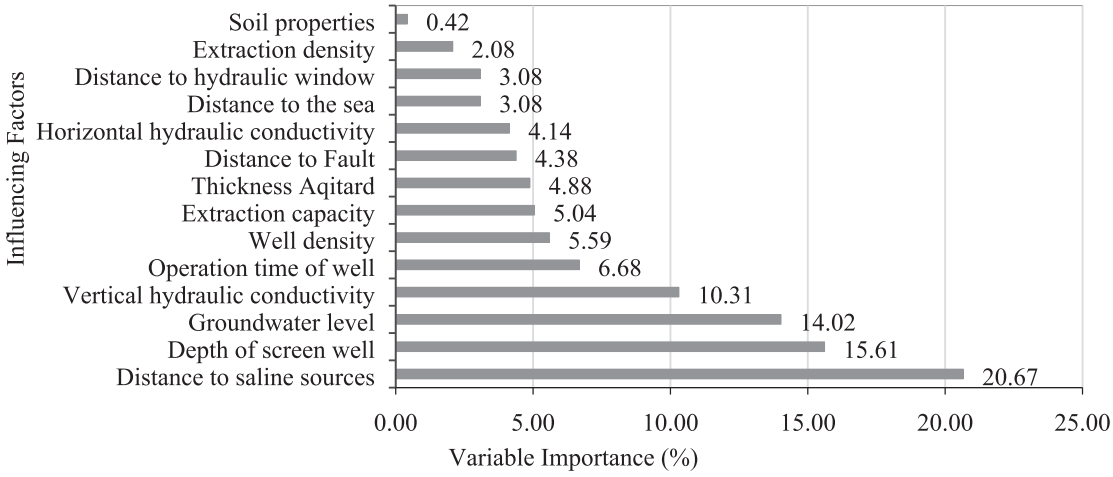

Fig. 5. Variable importance (%) of influencing factors for groundwater salinity prediction.

depletion. Regions severely affected by saltwater include the coastal estuary, the land along the river, and the central areas where the chloride concentrations in wells elevate to over 2,000 mg/L. It is noted that low chloride concentrations (Cl^- < 250 mg/L) in groundwater are predicted in coastal areas, even the production wells being located near the shoreline (<2 km) and being at −10.5 m below the mean sea level (m.a.m.s.l). This prediction result is also consistent with the observed data reported in other studies (Hoang and Bäumle, 2019; Tran et al., 2020). This could be explained by the geological and hydrogeological features in the VMD in which deep aquifers are covered by thick clay layers extending far from the shoreline (Hung Van et al., 2019). Similar observations were found in many coastal lowland areas in the world, such as in China (Han and Currell, 2018), in Thailand (Xiong et al., 2020), in India (Dhakate et al., 2020; Kumar et al., 2020), in Bangladesh (Datta et al., 2020; Seddique et al., 2019; Xiong et al., 2020), in Mexico (Mahlknecht et al., 2017; Mora et al., 2020), in Tunisia (Souid et al., 2018; Telahigue et al., 2018), in Brazil (Cary et al., 2015; Gomes et al., 2019), in the EU (Giménez-Forcada, 2014; Kazakis et al., 2016; Telahigue et al., 2020), and North America (Bond and Bredehoeft, 1987; Goebel et al., 2017). In contrast, in the central areas - far from the Sea (approximately 40 km inland), relatively high chloride concentration was predicted to be consistent with measured chloride concentration at the same locations. In these cases, salinity sources were trapped in these areas a long time ago due to rapid sea-level rise during the Holocene period (Delsman et al., 2014; Hung Van et al., 2019; Larsen et al., 2017). Besides, different density between saline groundwater at shallow aquifers and freshwater in deep aquifers often causes the movement of salinity downward (Ma et al., 2015). However, this process depends on

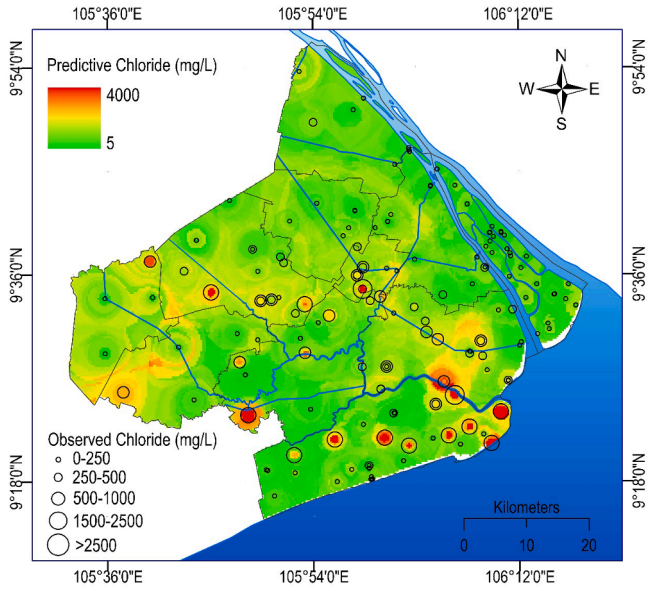

Fig. 6. Prediction of chloride concentration (mg/L) in the study area by using the CBR model.

the concentration of salinity at specific locations, the thickness of the aquitard, vertical and horizontal hydraulic conductivity, which has been illustrated by variable importance assessment.

The CBR model provides a prediction of salinity-affected area (Table 4) which the extreme salinity (>1000 mg/L) area is 43.23 km^2 (1.31%) while the moderate salinity (500 < Cl^- < 1,000 mg/L) area is 285.25 km^2 (8.61%). In addition, the prediction result shows that moderate to high salinity-affected areas are distributed heterogeneously in the study area. For example, profoundly salinity-affected sites were found not only along the coast but also in the central part of the province (Fig. 6), indicating complex groundwater salinization processes in the accumulative impacts of different factors, as found in Fig. 5.

It is noted that approximately 1,632.21 km^2 (49.28%) have low chloride concentration (Cl^- < 250 mg/L) located far from the paleo-saline sources and high groundwater level. This indicates that these influencing factors play an essential role in increasing chloride concentrations in groundwater. Our findings are in-line with recent studies (An et al., 2018; Nam et al., 2019), demonstrating the capability of the CBR model in predicting chloride concentration in coastal regions.

4.4. Estimation of the salinity-affected population

Population in the study area distributed unevenly, urban areas in the central and coastal regions have high population density with a maximum of 6,120 people/km^2. High population density distributes in central and the coastal regions may increase risks of groundwater salinity to water users because high salinity concentration was observed in these areas (Fig. 7).

As shown in Table 4, approximately half of the population in the study area (629,360 people) was using inadequate groundwater sources with salinity higher than the limit (<250 mg/L) of the WHO drinking water standard. Meanwhile, nearly 10% of the total population (124,123 people) might be influenced by water with moderate to high salinity concentration (500–2,500 mg/L). It was noted that although very high salinity (>2,500 mg/L) is located in a small area (2.82 km^2), it may significantly influence a large number of people due to massive groundwater exploitation in dense population areas such as central and coastal cities. Moreover, the number of people facing freshwater shortages are likely to increase shortly because of climate change-induced sea-level rise.

Table 4
Predictive results of affected area and population following five classes of salinity concentration in groundwater.

Magnitude	Very low	Low	Moderate	High	Very high
Salinity (mg/L)	<250	250–500	500–1000	1000–2500	>2500
Area (km2)	1632.21	1348.5	285.25	43.23	2.82
In percentage (%)	49.28%	40.72%	8.61%	1.31%	0.09%
People	670,640	505,237	106,872	16,195	1055
In percentage (%)	51.59%	38.86%	8.22%	1.25%	0.08%

5. Concluding remarks

This study is the first attempt to evaluate and compare the four *state-of-the-art* ML algorithms (RFR, XGBR, CBR, and LGBR) for accurately estimating groundwater salinity in the coastal multi-aquifers of the MD. The novel ML algorithm, namely, the CBR model along with the three well-known ML algorithms (LGBR, XGBR, and RFR) were trained and tested with the 14 influencing factors from the coastal multi-aquifers in the Mekong Delta, Vietnam. The results showed that the CBR model achieved satisfactory accuracy in the testing set (RMSE = 205.96, R^2 = 0.84) for estimating the salinity concentrations in coastal aquifers of the Mekong Delta. The CBR model offered the best accurate prediction and stability among the four ML algorithms, while the LGBR model had the lowest accuracy.

A more accurate prediction obtained by using the optimal features through the variable importance selection function in the CBR algorithm shows a similar result for training (RMSE = 26.51, R^2 = 0.99) and testing (RMSE = 193.51, R^2 = 0.86) with ten influencing factors, including the distance to saline sources, the depth of screen well, the groundwater level, the vertical hydraulic conductivity, the operation time, the well density, the extraction capacity, the thickness of aquitard, the distance to fault, and the horizontal hydraulic conductivity.

The groundwater salinity generalized from the CBR model is consistent with the measured data. Spatial distribution of salinity has a good correlation to the distance to saline sources, the depth of screen well, the groundwater level, and the vertical hydraulic conductivity. This indicates that the risk of groundwater salinization in the study area

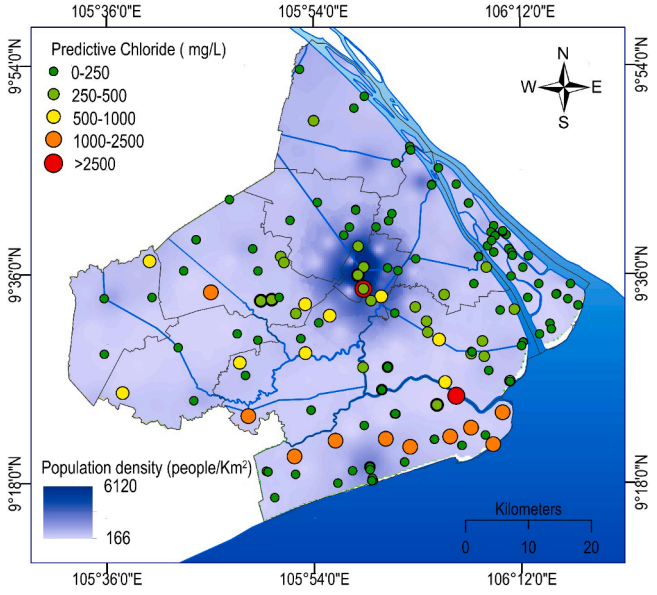

Fig. 7. Spatial distribution of predicted Chloride and population density in the study area.

depends significantly on both the pumping activities and the aquifer properties. Overall, as an ensemble-based boosting algorithm, the CBR technique made significant improvement in overall accuracy, stability and computational cost compared to those of the RFR, the XGBR and the LGBR models. Therefore, we conclude that the CBR model shows the high potential applications for salinity estimation in coastal multi-aquifers of the Mekong Delta, Vietnam. The effectiveness of the proposed model developed in this work should be tested and compared in other parts of the world with similar geographic and hydrographical conditions.

CRediT authorship contribution statement

Dang An Tran: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Maki Tsujimura:** Supervision, Writing - review & editing. **Nam Thang Ha:** Conceptualization, Methodology, Software, Formal analysis, Writing - review & editing. **Van Tam Nguyen:** Conceptualization, Methodology, Software, Formal analysis, Writing - review & editing. **Doan Van Binh:** Conceptualization, Methodology, Software, Formal analysis, Writing - review & editing. **Thanh Duc Dang:** Conceptualization, Methodology, Software, Formal analysis, Writing - review & editing. **Quang-Van Doan:** Conceptualization, Methodology, Software, Formal analysis, Writing - review & editing. **Dieu Tien Bui:** Conceptualization, Methodology, Software, Formal analysis, Writing - review & editing. **Trieu Anh Ngoc:** Conceptualization, Methodology, Software, Formal analysis, Writing - review & editing. **Le Vo Phu:** Conceptualization, Methodology, Software, Formal analysis, Writing - review & editing. **Pham Thi Bich Thuc:** Conceptualization, Methodology, Software, Formal analysis, Writing - review & editing. **Tien Dat Pham:** Conceptualization, Methodology, Software, Formal analysis, Writing - review & editing.

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.

References

Abdelhamid, H., Javadi, A., Abd-Elaty, I., Sherif, M., 2016. Simulation of seawater intrusion in the Nile Delta aquifer under the conditions of climate change.

Aeschbach-Hertig, W., Gleeson, T., 2012. Regional strategies for the accelerating global problem of groundwater depletion. Nat. Geosci. 5, 853–861.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19, 716–723.

Akbari, M., Najafi Alamdarlo, H., Mosavi, S.H., 2020. The effects of climate change and groundwater salinity on farmers' income risk. Ecol. Ind. 110, 105893.

Akter, S., 2019. Impact of drinking water salinity on children's education: empirical evidence from coastal Bangladesh. Sci. Total Environ. 690, 1331–1341.

An, T.D., Tsujimura, M., Phu, V.L., Ha, D.T., Hai, N.V., 2018. Isotopic and Hydrogeochemical Signatures in Evaluating Groundwater Quality in the Coastal Area of the Mekong Delta, Vietnam. In: Tien Bui, D., Ngoc Do, A., Bui, H.-B., Hoang, N.-D. (Eds.), Advances and Applications in Geospatial Technology and Earth Resources. Springer International Publishing, Cham, pp. 293–314.

Badaruddin, S., Werner, A.D., Morgan, L.K., 2017. Characteristics of active seawater intrusion. J. Hydrol. 551, 632–647.

Behera, A.K., Chakrapani, G.J., Kumar, S., Rai, N., 2019. Identification of seawater intrusion signatures through geochemical evolution of groundwater: a case study based on coastal region of the Mahanadi delta, Bay of Bengal, India. Nat. Hazards 97, 1209–1230.

Binh, D.V., Kantoush, S., Sumi, T., 2020. Changes to long-term discharge and sediment loads in the Vietnamese Mekong Delta caused by upstream dams. Geomorphology 353, 107011.

Blasco, M., Auqué, L.F., Gimeno, M.J., 2019. Geochemical evolution of thermal waters in carbonate – evaporitic systems: the triggering effect of halite dissolution in the dedolomitisation and albitisation processes. J. Hydrol. 570, 623–636.

Boluda-Botella, N., Gomis-Yagües, V., Ruiz-Beviá, F., 2008. Influence of transport parameters and chemical properties of the sediment in experiments to measure reactive transport in seawater intrusion. J. Hydrol. 357, 29–41.

Bond, L.D., Bredehoeft, J.D., 1987. Origins of seawater intrusion in a coastal aquifer — a case study of the Pajaro Valley, California. J. Hydrol. 92, 363–388.

Braden, B., 1986. The Surveyor's Area Formula. College Math. J. 17, 326–337.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Burnham, K.P., Anderson, D.R., 2004. Multimodel Inference: understanding AIC and BIC in Model Selection. Soc. Methods Res. 33, 261–304.

Carretero, S., Rapaglia, J., Bokuniewicz, H., Kruse, E., 2013. Impact of sea-level rise on saltwater intrusion length into the coastal aquifer, Partido de La Costa, Argentina. Cont. Shelf Res. 61–62, 62–70.

Cary, L., Petelet-Giraud, E., Bertrand, G., Kloppmann, W., Aquilina, L., Martins, V., Hirata, R., Montenegro, S., Pauwels, H., Chatton, E., Franzen, M., Aurouet, A., Lasseur, E., Picot, G., Guerrot, C., Fléhoc, C., Labasque, T., Santos, J.G., Paiva, A., Braibant, G., Pierre, D., 2015. Origins and processes of groundwater salinization in the urban coastal aquifers of Recife (Pernambuco, Brazil): a multi-isotope approach. Sci. Total Environ. 530–531, 411–429.

Chatton, E., Aquilina, L., Pételet-Giraud, E., Cary, L., Bertrand, G., Labasque, T., Hirata, R., Martins, V., Montenegro, S., Vergnaud, V., Aurouet, A., Kloppmann, W., Pauwels,, 2016. Glacial recharge, salinisation and anthropogenic contamination in the coastal aquifers of Recife (Brazil). Sci. Environ. 569-570, 1114–1125.

Chekirbane, A., Tsujimura, M., Kawachi, A., Lachaal, F., Isoda, H., Tarhouni, J., 2014. Use of a time-domain electromagnetic method with geochemical tracers to explore the salinity anomalies in a small coastal aquifer in north-eastern Tunisia. Hydrogeol. J. 22, 1777–1794.

Chen, C., He, W., Zhou, H., Xue, Y., Zhu, M., 2020. A comparative study among machine learning and numerical models for simulating groundwater dynamics in the Heihe River Basin, northwestern China. Sci. Rep. 10, 3904.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, San Francisco, California, USA, pp. 785–794.

Chen, Z.-Y., Zhang, T.-H., Zhang, R., Zhu, Z.-M., Yang, J., Chen, P.-Y., Ou, C.-Q., Guo, Y., 2019. Extreme gradient boosting model to estimate PM2.5 concentrations with missing-filled satellite data in China. Atmos. Environ. 202, 180–189.

Crestani, E., Camporese, M., Salandin, P., 2019. Technical note: an alternative approach to laboratory benchmarking of saltwater intrusion in coastal aquifers. Hydrol. Earth Syst. Sci. Discuss. 2019, 1–18.

Criminisi, A., 2011. Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Found. Trends® in Comput. Graphics Vision 7, 81–227.

Dang, T.D., Cochrane, T.A., Arias, M.E., Tri, V.P.D., 2018. Future hydrological alterations in the Mekong Delta under the impact of water resources development, land subsidence and sea level rise. J. Hydrol.: Reg. Stud. 15, 119–133.

Datta, D.K., Ghosh, P.K., Karim, M.R., Rahman, M.M., 2020. Geochemical options for water security in a coastal urban agglomerate of Lower Bengal Delta, Bangladesh. J. Geochem. Explor. 209, 106440.

Delsman, J.R., Hu-a-ng, K.R.M., Vos, P.C., de Louw, P.G.B., Oude Essink, G.H.P., Stuyfzand, P.J., Bierkens, M.F.P., 2014. Paleo-modeling of coastal saltwater intrusion during the Holocene: an application to the Netherlands. Hydrol. Earth Syst. Sci. 18, 3891–3905.

Dev, V.A., Eden, M.R., 2019. Formation lithology classification using scalable gradient boosted decision trees. Comput. Chem. Eng. 128, 392–404.

Dhakate, R., Ratnalu, G.V., Sankaran, S., 2020. Hydrogeochemical and isotopic study for evaluation of seawater intrusion into shallow coastal aquifers of Udupi District, Karnataka, India. Geochemistry 125647.

Dorogush, A.V., Ershov, V., Gulin, A., 2018. CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363.

Elmahdy, S.I., Mohamed, M.M., 2013. Influence of geological structures on groundwater accumulation and groundwater salinity in Musandam Peninsula, UAE and Oman. Geocarto Int. 28, 453–472.

Essaid, H.I., Caldwell, R.R., 2017. Evaluating the impact of irrigation on surface water – groundwater interaction and stream temperature in an agricultural watershed. Sci. Total Environ. 599–600, 581–596.

Famiglietti, J.S., 2014. The global groundwater crisis. Nat. Clim. Change 4, 945–948.

Fan, J., Ma, X., Wu, L., Zhang, F., Yu, X., Zeng, W., 2019. Light gradient boosting machine: an efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. Agric. Water Manag. 225, 105758.

Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., Lu, X., Xiang, Y., 2018. Comparison of support vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: a case study in China. Energy Convers. Manage. 164, 102–111.

Ferguson, G., Gleeson, T., 2012. Vulnerability of coastal aquifers to groundwater use and climate change. Nat. Clim. Change 2, 342–345.

Foster, S., Pulido-Bosch, A., Vallejos, Á., Molina, L., Llop, A., MacDonald, A.M., 2018. Impact of irrigated agriculture on groundwater-recharge salinity: a major sustainability concern in semi-arid regions. Hydrogeol. J. 26, 2781–2791.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232.

Gejl, R.N., Bjerg, P.L., Henriksen, H.J., Bitsch, K., Troldborg, L., Schullehner, J., Rasmussen, J., Rygaard, M., 2020. Relating wellfield drawdown and water quality to aquifer sustainability – a method for assessing safe groundwater abstraction. Ecol. Ind. 110, 105782.

Giménez-Forcada, E., 2014. Space/time development of seawater intrusion: a study case in Vinaroz coastal plain (Eastern Spain) using HFE-Diagram, and spatial distribution of hydrochemical facies. J. Hydrol. 517, 617–627.

Goebel, M., Pidlisecky, A., Knight, R., 2017. Resistivity imaging reveals complex pattern of saltwater intrusion along Monterey coast. J. Hydrol. 551, 746–755.

Gomes, O.V.O., Marques, E.D., Kütter, V.T., Aires, J.R., Travi, Y., Silva-Filho, E.V., 2019. Origin of salinity and hydrogeochemical features of porous aquifers from

northeastern Guanabara Bay, Rio de Janeiro, SE - Brazil. J. Hydrol.: Reg. Stud. 22, 100601.
Gong, M., Bai, Y., Qin, J., Wang, J., Yang, P., Wang, S., 2020. Gradient boosting machine for predicting return temperature of district heating system: a case study for residential buildings in Tianjin. J. Build. Eng. 27, 100950.
Gossel, W., Sefelnasr, A., Wycisk, P., 2010. Modelling of paleo-saltwater intrusion in the northern part of the Nubian Aquifer System, Northeast Africa. Hydrogeol. J. 18, 1447–1463.
Guo, Q., Huang, J., Zhou, Z., Wang, J., 2019. Experiment and numerical simulation of seawater intrusion under the influences of tidal fluctuation and groundwater exploitation in coastal multilayered aquifers. Geofluids 2019, 2316271.
Ha, T.P., Dieperink, C., Dang Tri, V.P., Otter, H.S., Hoekstra, P., 2018. Governance conditions for adaptive freshwater management in the Vietnamese Mekong Delta. J. Hydrol. 557, 116–127.
Hamer, T., Dieperink, C., Tri, V.P.D., Otter, H.S., Hoekstra, P., 2020. The rationality of groundwater governance in the Vietnamese Mekong Delta's coastal zone. Int. J. Water Resour. Dev. 36, 127–148.
Han, D., Currell, M.J., 2018. Delineating multiple salinization processes in a coastal plain aquifer, northern China: hydrochemical and isotopic evidence. Hydrol. Earth Syst. Sci. 22, 3473–3491.
Han, D., Post, V.E.A., Song, X., 2015. Groundwater salinization processes and reversibility of seawater intrusion in coastal carbonate aquifers. J. Hydrol. 531, 1067–1080.
Hoang, H.T., Bäumle, R., 2019. Complex hydrochemical characteristics of the Middle-Upper Pleistocene aquifer in Soc Trang Province, Southern Vietnam. Environ. Geochem. Health 41, 325–341.
Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., Zeng, W., Zhou, H., 2019. Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. J. Hydrol. 574, 1029–1041.
Hung Van, P., Van Geer, F.C., Bui Tran, V., Dubelaar, W., Oude Essink, G.H.P., 2019. Paleo-hydrogeological reconstruction of the fresh-saline groundwater distribution in the Vietnamese Mekong Delta since the late Pleistocene. J. Hydrol.: Reg. Stud. 23, 100594.
Johnson, N.E., Bonczak, B., Kontokosta, C.E., 2018. Using a gradient boosting model to improve the performance of low-cost aerosol monitors in a dense, heterogeneous urban environment. Atmos. Environ. 184, 9–16.
Kagabu, M., Shimada, J., Delinom, R., Tsujimura, M., Taniguchi, M., 2011. Groundwater flow system under a rapidly urbanizing coastal city as determined by hydrogeochemistry. J. Asian Earth Sci. 40, 226–239.
Kanagaraj, G., Elango, L., Sridhar, S.G.D., Gowrisankar, G., 2018. Hydrogeochemical processes and influence of seawater intrusion in coastal aquifers south of Chennai, Tamil Nadu, India. Environ. Sci. Pollut. Res. 25, 8989–9011.
Kang, P., Lin, Z., Teng, S., Zhang, G., Guo, L., Zhang, W., 2019. Catboost-based framework with additional user information for social media popularity prediction. In: Proceedings of the 27th ACM International Conference on Multimedia. Association for Computing Machinery, Nice, France, pp. 2677–2681.
Kaur, L., Rishi, M.S., Singh, G., Nath Thakur, S., 2020. Groundwater potential assessment of an alluvial aquifer in Yamuna sub-basin (Panipat region) using remote sensing and GIS techniques in conjunction with analytical hierarchy process (AHP) and catastrophe theory (CT). Ecol. Ind. 110, 105850.
Kazakis, N., Pavlou, A., Vargemezis, G., Voudouris, K.S., Soulios, G., Pliakas, F., Tsokas, G., 2016. Seawater intrusion mapping using electrical resistivity tomography and hydrochemical data. An application in the coastal area of eastern Thermaikos Gulf, Greece. Sci. Total Environ. 543, 373–387.
Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. Lightgbm: a highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems, pp. 3146–3154.
Khaska, M., Le Gal La Salle, C., Lancelot, J., team, A., Mohamad, A., Verdoux, P., Noret, A., Simler, R., 2013. Origin of groundwater salinity (current seawater vs. saline deep water) in a coastal karst aquifer based on Sr and Cl isotopes. Case study of the La Clape massif (southern France). Appl. Geochem. 37, 212–227.
Kopsiaftis, G., Protopapadakis, E., Voulodimos, A., Doulamis, N., Mantoglou, A., 2019. Gaussian process regression tuned by bayesian optimization for seawater intrusion prediction. Comput. Intell Neurosci. 2859429-2859429.
Korres, N.E., Varanasi, V.K., Slaton, N.A., Price, A.J., Bararpour, T., 2019. Chapter 8 - Effects of Salinity on Rice and Rice Weeds: Short- and Long-Term Adaptation Strategies and Weed Management. In: Hasanuzzaman, M., Fujita, M., Nahar, K., Biswas, J.K. (Eds.), Advances in Rice Research for Abiotic Stress Tolerance. Woodhead Publishing, pp. 159–176.
Kumar, P.J.S., Jegathambal, P., Babu, B., Kokkat, A., James, E.J., 2020. A hydrogeochemical appraisal and multivariate statistical analysis of seawater intrusion in point calimere wetland, lower Cauvery region, India. Groundwater Sustainable Dev. 11, 100392.
Lal, A., Datta, B., 2019. Multi-objective groundwater management strategy under uncertainties for sustainable control of saltwater intrusion: solution for an island country in the South Pacific. J. Environ. Manage. 234, 115–130.
Langenheim, N., White, M., Barton, J., Eagleson, S., 2017. Designing with data for urban resilience. In: Geertman, S., Allan, A., Pettit, C., Stillwell, J. (Eds.), Planning Support Science for Smarter Urban Futures. Springer International Publishing, Cham, pp. 113–133.
Lapworth, D.J., Krishan, G., MacDonald, A.M., Rao, M.S., 2017. Groundwater quality in the alluvial aquifer system of northwest India: new evidence of the extent of anthropogenic and geogenic contamination. Sci. Total Environ. 599–600, 1433–1444.
Larsen, F., Tran, L.V., Van Hoang, H., Tran, L.T., Christiansen, A.V., Pham, N.Q., 2017. Groundwater salinity influenced by Holocene seawater trapped in incised valleys in the Red River delta plain. Nat. Geosci. 10, 376–381.
Li, B., Yang, G., Wan, R., Dai, X., Zhang, Y., 2016. Comparison of random forests and other statistical methods for the prediction of lake water level: a case study of the Poyang Lake in China. Hydrol. Res.
Li, J., Lu, W., Wang, H., Fan, Y., Chang, Z., 2020. Groundwater contamination source identification based on a hybrid particle swarm optimization-extreme learning machine. J. Hydrol. 584, 124657.
Lim, S., Chi, S., 2019. Xgboost application on bridge management systems for proactive damage estimation. Adv. Eng. Inf. 41, 100922.
Liu, D., Liu, C., Fu, Q., Li, T., Imran, K.M., Cui, S., Abrar, F.M., 2017. ELM evaluation model of regional groundwater quality based on the crow search algorithm. Ecol. Ind. 81, 302–314.
Liudmila Prokhorenkova, G.G., Aleksandr Vorobev, Anna Veronika Dorogush, Andrey Gulin, 2017. CatBoost: unbiased boosting with categorical features. arXiv preprint, 1706.09516v5.
Ma, Q., Li, H., Wang, X., Wang, C., Wan, L., Wang, X., Jiang, X., 2015. Estimation of seawater–groundwater exchange rate: case study in a tidal flat with a large-scale seepage face (Laizhou Bay, China). Hydrogeol. J. 23, 265–275.
Ma, Y., Liu, Z.-H., Xi, B.-D., He, X.-S., Li, Q.-L., Qi, Y.-J., Jin, M.-Y., Guo, Y., 2019. Characteristics of groundwater pollution in a vegetable cultivation area of typical facility agriculture in a developed city. Ecol. Ind. 105, 709–716.
Mahlknecht, J., Merchán, D., Rosner, M., Meixner, A., Ledesma-Ruiz, R., 2017. Assessing seawater intrusion in an arid coastal aquifer under high anthropogenic influence using major constituents, Sr and B isotopes in groundwater. Sci. Total Environ. 587–588, 282–295.
Mahmoodzadeh, D., Karamouz, M., 2019. Seawater intrusion in heterogeneous coastal aquifers under flooding events. J. Hydrol. 568, 1118–1130.
Matsuzaka, Y., Hosaka, T., Ogaito, A., Yoshinari, K., Uesawa, Y., 2020. Prediction model of aryl hydrocarbon receptor activation by a novel QSAR approach, deepsnap–deep learning. Molecules 25.
Minderhoud, P.S.J., Coumou, L., Erban, L.E., Middelkoop, H., Stouthamer, E., Addink, E. A., 2018. The relation between land use and subsidence in the Vietnamese Mekong delta. Sci. Total Environ. 634, 715–726.
Minderhoud, P.S.J., Erkens, G., Pham, V.H., Bui, V.T., Erban, L., Kooi, H., Stouthamer, E., 2017. Impacts of 25 years of groundwater extraction on subsidence in the Mekong delta, Vietnam. Environ. Res. Lett. 12, 064006.
Moazamnia, M., Hassanzadeh, Y., Nadiri, A.A., Sadeghfam, S., 2020. Vulnerability indexing to saltwater intrusion from models at two levels using artificial intelligence multiple model (AIMM). J. Environ. Manage. 255, 109871.
Moghaddam, D.D., Rahmati, O., Panahi, M., Tiefenbacher, J., Darabi, H., Haghizadeh, A., Haghighi, A.T., Nalivan, O.A., Tien Bui, D., 2020. The effect of sample size on different machine learning models for groundwater potential mapping in mountain bedrock aquifers. CATENA 187, 104421.
Mohanty, A.K., Rao, V.V.S.G., 2019. Hydrogeochemical, seawater intrusion and oxygen isotope studies on a coastal region in the Puri District of Odisha, India. CATENA 172, 558–571.
Mora, A., Mahlknecht, J., Ledesma-Ruiz, R., Sanford, W.E., Lesser, L.E., 2020. Dynamics of major and trace elements during seawater intrusion in a coastal sedimentary aquifer impacted by anthropogenic activities. J. Contam. Hydrol. 232, 103653.
Nahian, M.A., Ahmed, A., Lázár, A.N., Hutton, C.W., Salehin, M., Streatfield, P.K., 2018. Drinking water salinity associated health crisis in coastal Bangladesh. Elem. Sci. Anth. 6, 14.
Nam, N.D.G., Akira, G., Kazutoshi, O., Trung, N.H., Ngan, N.V.C., 2019. Assessment of groundwater quality and its suitability for domestic and irrigation use in the coastal zone of the Mekong Delta, Vietnam. In: Stewart, M.A., Coclanis, P.A. (Eds.), Water and Power: Environmental Governance and Strategies for Sustainability in the Lower Mekong Basin. Springer International Publishing, Cham, pp. 173–185.
Naser, A.M., Rahman, M., Unicomb, L., Doza, S., Ahmed, K.M., Uddin, M.N., Selim, S., Gribble, M.O., Anand, S., Clasen, T.F., Luby, S.P., 2017. Drinking water salinity and kidney health in southwest coastal Bangladesh: baseline findings of a community-based stepped-wedge randomised trial. The Lancet 389, S15.
Ni, L., Wang, D., Wu, J., Wang, Y., Tao, Y., Zhang, J., Liu, J., 2020. Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. J. Hydrol. 586, 124901.
Pan, Y., Chen, S., Qiao, F., Ukkusuri, S.V., Tang, K., 2019. Estimation of real-driving emissions for buses fueled with liquefied natural gas based on gradient boosted regression trees. Sci. Total Environ. 660, 741–750.
Park, J., Kwock, C.K., 2015. Sodium intake and prevalence of hypertension, coronary heart disease, and stroke in Korean adults. J. Ethnic Foods 2, 92–96.
Parvin, S., Van Geel, M., Yeasmin, T., Lievens, B., Honnay, O., 2019. Variation in arbuscular mycorrhizal fungal communities associated with lowland rice (Oryza sativa) along a gradient of soil salinity and arsenic contamination in Bangladesh. Sci. Total Environ. 686, 546–554.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., Louppe, G., 2012. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12.
Pham, D.T., Yokoya, N., Xia, J., Ha, T.N., Le, N.N., Nguyen, T.T., Dao, H.T., Vu, T.T., Pham, D.T., Takeuchi, W., 2020a. Comparison of machine learning methods for estimating mangrove above-ground biomass using multiple source remote sensing data in the Red River Delta Biosphere Reserve, Vietnam. Remote Sens. 12.
Pham, H.V., Tsai, F.T.C., 2016. Optimal observation network design for conceptual model discrimination and uncertainty reduction. Water Resour. Res. 52, 1245–1264.

Pham, T.D., Le, N.N., Ha, N.T., Nguyen, L.V., Xia, J., Yokoya, N., To, T.T., Trinh, H.X., Kieu, L.Q., Takeuchi, W., 2020b. Estimating Mangrove above-ground biomass using extreme gradient boosting decision trees algorithm with fused Sentinel-2 and ALOS-2 PALSAR-2 data in can gio biosphere reserve, Vietnam. Remote Sens. 12, 777.
Pham, T.D., Yoshino, K., Bui, D.T., 2017. Biomass estimation of Sonneratia caseolaris (l.) Engler at a coastal area of Hai Phong city (Vietnam) using ALOS-2 PALSAR imagery and GIS-based multi-layer perceptron neural networks. GIScience Remote Sens. 54, 329–353.
Podgorski, J., Labhasetwar, P., Saha, D., Berg, M., 2018. Prediction modeling and mapping of groundwater fluoride contamination throughout India. Environ. Sci. Technol.
Pourghasemi, H.R., Sadhasivam, N., Yousefi, S., Tavangar, S., Ghaffari Nazarlou, H., Santosh, M., 2020. Using machine learning algorithms to map the groundwater recharge potential zones. J. Environ. Manage. 265, 110525.
Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features. Adv. Neural Inf. Process. Syst. 6638–6648.
Radanielson, A.M., Gaydon, D.S., Li, T., Angeles, O., Roth, C.H., 2018. Modeling salinity effect on rice growth and grain yield with ORYZA v3 and APSIM-Oryza. Eur. J. Agron. 100, 44–55.
Rahaman, M.A., Rahman, M.M., Nazimuzzaman, M., 2020. Impact of Salinity on Infectious Disease Outbreaks: experiences from the Global Coastal Region. In: Leal Filho, W., Wall, T., Azul, A.M., Brandli, L., Özuyar, P.G. (Eds.), Good Health and Well-Being. Springer International Publishing, Cham, pp. 415–424.
Rahman, A.T.M.S., Hosono, T., Quilty, J.M., Das, J., Basak, A., 2020a. Multiscale groundwater level forecasting: coupling new machine learning approaches with wavelet transforms. Adv. Water Resour. 141, 103595.
Rahman, S., Irfan, M., Raza, M., Moyeezullah Ghori, K., Yaqoob, S., Awais, M., 2020b. Performance analysis of boosting classifiers in recognizing activities of daily living. Int. J. Environ. Res. Public Health 17.
Rahmati, O., Choubin, B., Fathabadi, A., Coulon, F., Soltani, E., Shahabi, H., Mollaefar, E., Tiefenbacher, J., Cipullo, S., Ahmad, B.B., Tien Bui, D., 2019. Predicting uncertainty of machine learning models for modelling nitrate pollution of groundwater using quantile regression and UNEEC methods. Sci. Total Environ. 688, 855–866.
Rajaee, T., Ebrahimi, H., Nourani, V., 2019. A review of the artificial intelligence methods in groundwater level modeling. J. Hydrol. 572, 336–351.
Ransom, K.M., Nolan, B.T., Traum, J.A., Faunt, C.C., Bell, A.M., Gronberg, J.A.M., Wheeler, D.C., Rosecrans, C.Z., Jurgens, B., Schwarz, G.E., Belitz, K., Eberts, S.M., Kourakos, G., Harter, T., 2017. A hybrid machine learning model to predict and visualize nitrate concentration throughout the Central Valley aquifer, California, USA. Sci. Total Environ. 601-602, 1160–1172.
Robinson, G., Moutari, S., Ahmed, A.A., Hamill, G.A., 2018. An advanced calibration method for image analysis in laboratory-scale seawater intrusion problems. Water Resour. Manage. 32, 3087–3102.
Sahoo, S., Russo, T.A., Elliott, J., Foster, I., 2017. Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S. Water Resour. Res. 53, 3878–3895.
Sajedi-Hosseini, F., Malekian, A., Choubin, B., Rahmati, O., Cipullo, S., Coulon, F., Pradhan, B., 2018. A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. Sci. Total Environ. 644, 954–962.
Scharping, R.J., Garman, K.M., Henry, R.P., Eswara, P.J., Garey, J.R., 2018. The fate of urban springs: pumping-induced seawater intrusion in a phreatic cave. J. Hydrol. 564, 230–245.
Seddique, A.A., Masuda, H., Anma, R., Bhattacharya, P., Yokoo, Y., Shimizu, Y., 2019. Hydrogeochemical and isotopic signatures for the identification of seawater intrusion in the paleobeach aquifer of Cox's Bazar city and its surrounding area, south-east Bangladesh. Groundwater Sustainable Dev. 9, 100215.
Shammi, M., Rahman, M.M., Bondad, S.E., Bodrud-Doza, M., 2019. Impacts of salinity intrusion in community health: a review of experiences on drinking water sodium from coastal areas of bangladesh. Healthcare (Basel) 7, 50.
Shi, W., Lu, C., Werner, A.D., 2020. Assessment of the impact of sea-level rise on seawater intrusion in sloping confined coastal aquifers. J. Hydrol. 586, 124872.
Shrestha, S., Bach, T.V., Pandey, V.P., 2016. Climate change impacts on groundwater resources in Mekong Delta under representative concentration pathways (RCPs) scenarios. Environ. Sci. Policy 61, 1–13.
Singh, A., 2015. Managing the environmental problem of seawater intrusion in coastal aquifers through simulation–optimization modeling. Ecol. Ind. 48, 498–504.
Singh, L.K., Jha, M.K., Chowdary, V.M., 2018. Assessing the accuracy of GIS-based Multi-Criteria Decision Analysis approaches for mapping groundwater potential. Ecol. Ind. 91, 24–37.
Smajgl, A., Toan, T.Q., Nhan, D.K., Ward, J., Trung, N.H., Tri, L.Q., Tri, V.P.D., Vu, P.T., 2015. Responding to rising sea levels in the Mekong Delta. Nat. Clim. Change 5, 167–174.
Souid, F., Agoubi, B., Telahigue, F., Chahlaoui, A., Kharroubi, A., 2018. Groundwater salinization and seawater intrusion tracing based on Lithium concentration in the shallow aquifer of Jerba Island, southeastern Tunisia. J. Afr. Earth Sc. 138, 233–246.
Stein, S., Yechieli, Y., Shalev, E., Kasher, R., Sivan, O., 2019. The effect of pumping saline groundwater for desalination on the fresh–saline water interface dynamics. Water Res. 156, 46–57.
Stone, M., 1979. Comments on Model Selection Criteria of Akaike and Schwarz. J. R. Stat. Soc. Ser. B (Methodological) 41, 276–278.
Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res.: Atmos. 106, 7183–7192.
Telahigue, F., Agoubi, B., Souid, F., Kharroubi, A., 2018. Assessment of seawater intrusion in an arid coastal aquifer, south-eastern Tunisia, using multivariate statistical analysis and chloride mass balance. Phys. Chem. Earth, Parts A/B/C 106, 37–46.
Telahigue, F., Mejri, H., Mansouri, B., Souid, F., Agoubi, B., Chahlaoui, A., Kharroubi, A., 2020. Assessing seawater intrusion in arid and semi-arid Mediterranean coastal aquifers using geochemical approaches. Phys. Chem. Earth, Parts A/B/C 115, 102811.
Tien Bui, D., Hoang, N.-D., Nhu, V.-H., 2018. A swarm intelligence-based machine learning approach for predicting soil shear strength for road construction: a case study at Trung Luong National Expressway Project (Vietnam). Eng. Comput.
Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., 2012. Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and Naïve Bayes models. Math. Probl. Eng. 2012, 974638.
Tien Bui, D., Tuan, T.A., Klempe, H., Pradhan, B., Revhaug, I., 2016. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides 13, 361–378.
Tran, D.A., Tsujimura, M., Vo, L.P., Nguyen, V.T., Kambuku, D., Dang, T.D., 2020. Hydrogeochemical characteristics of a multi-layered coastal aquifer system in the Mekong Delta, Vietnam. Environ. Geochem. Health 42, 661–680.
Tran, D.A., Tsujimura, M., Vo, L.P., Nguyen, V.T., Nguyen, L.D., Dang, T.D., 2019. Stable isotope characteristics of water resources in the coastal area of the Vietnamese Mekong Delta. Isot. Environ. Health Stud. 55, 566–587.
Tweed, S., Celle-Jeanton, H., Cabot, L., Huneau, F., De Montety, V., Nicolau, N., Travi, Y., Babic, M., Aquilina, L., Vergnaud-Ayraud, V., Leblanc, M., 2018. Impact of irrigated agriculture on groundwater resources in a temperate humid region. Sci. Total Environ. 613–614, 1302–1316.
Velasco, J., Gutiérrez-Cánovas, C., Botella-Cruz, M., Sánchez-Fernández, D., Arribas, P., Carbonell, J.A., Millán, A., Pallarés, S., 2019. Effects of salinity changes on aquatic organisms in a multiple stressor context. Philos. Trans. R. Soc. B: Biol. Sci. 374, 20180011.
Vineis, P., Chan, Q., Khan, A., 2011. Climate change impacts on water salinity and health. J. Epidemiol. Global Health 1, 5–10.
Voss, C.I., Souza, W.R., 1987. Variable density flow and solute transport simulation of regional aquifers containing a narrow freshwater-saltwater transition zone. Water Resour. Res. 23, 1851–1866.
Vrieze, S.I., 2012. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Psychol. Methods 17, 228–243.
Wagner, F., Tran, V.B., Renaud, F.G., 2012. Groundwater resources in the mekong delta: availability, utilization and risks. In: Renaud, F.G., Kuenzer, C. (Eds.), The Mekong Delta System: Interdisciplinary Analyses of a River Delta. Springer, Netherlands, Dordrecht, pp. 201–220.
Walter, J., Chesnaux, R., Cloutier, V., Gaboury, D., 2017. The influence of water/rock – water/clay interactions and mixing in the salinization processes of groundwater. J. Hydrol.: Reg. Stud. 13, 168–188.
Walther, M., Stoeckl, L., Morgan, L.K., 2020. Post-pumping seawater intrusion at the field scale: Implications for coastal aquifer management. Adv. Water Resour. 138, 103561.
Wang, C.-M., Huang, Y.-F., 2009. Evolutionary-based feature selection approaches with new criteria for data mining: a case study of credit approval data. Expert Syst. Appl. 36, 5900–5908.
Winkel, L., Berg, M., Amini, M., Hug, S.J., Annette Johnson, C., 2008. Predicting groundwater arsenic contamination in Southeast Asia from surface parameters. Nat. Geosci. 1, 536–542.
Xiong, G.-Y., Chen, G.-Q., Xu, X.-Y., Liu, W.-Q., Fu, T.-F., Khokiattiwong, S., Kornkanitnan, N., Ali Seddique, A., Shi, X.-F., Liu, S.-F., Su, Q., Xu, X.-L., 2020. A comparative study on hydrochemical evolution and quality of groundwater in coastal areas of Thailand and Bangladesh. J. Asian Earth Sci. 195, 104336.
Yadav, B., Mathur, S., Ch, S., Yadav, B.K., 2018. Data-based modelling approach for variable density flow and solute transport simulation in a coastal aquifer. Hydrol. Sci. J. 63, 210–226.
Yechieli, Y., Yokochi, R., Zilberbrand, M., Lu, Z.-T., Purtschert, R., Sueltenfuss, J., Jiang, W., Zappala, J., Mueller, P., Bernier, R., Avrahamov, N., Adar, E., Talhami, F., Livshitz, Y., Burg, A., 2019. Recent seawater intrusion into deep aquifer determined by the radioactive noble-gas isotopes 81Kr and 39Ar. Earth Planet. Sci. Lett. 507, 21–29.
Yousefi, S., Avand, M., Yariyan, P., Pourghasemi, H.R., Keesstra, S., Tavangar, S., Tabibian, S., 2020. A novel GIS-based ensemble technique for rangeland downward trend mapping as an ecological indicator change. Ecol. Ind. 117, 106591.
Yu, X., Michael, H.A., 2019. Mechanisms, configuration typology, and vulnerability of pumping-induced seawater intrusion in heterogeneous aquifers. Adv. Water Resour. 128, 117–128.
Yu, X., Wang, Y., Wu, L., Chen, G., Wang, L., Qin, H., 2020. Comparison of support vector regression and extreme gradient boosting for decomposition-based data-driven 10-day streamflow forecasting. J. Hydrol. 582, 124293.
Yu, X., Xin, P., Lu, C., 2019. Seawater intrusion and retreat in tidally-affected unconfined aquifers: Laboratory experiments and numerical simulations. Adv. Water Resour. 132, 103393.
Zhang, Y., Zhao, Z., Zheng, J., 2020. CatBoost: a new approach for estimating daily reference crop evapotranspiration in arid and semi-arid regions of Northern China. J. Hydrol. 125087.