GROUNDWATER POTENTIAL MODEL USING STATISTICS AND HYBRID MACHINE

LEARNING APPROACHES

Mr. Ngoc Thanh Nguyen

A Dissertation Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in Environmental Science

Inter-Department of Environmental Science

GRADUATE SCHOOL

Chulalongkorn University

Academic Year 2022

แบบจำลองศักยภาพน้ำบาดาลโดยวิธีทางสถิติและการเรียนรู้ของเครื่องแบบไฮบริด

นายงอค ธานห์ เหวียน

Thesis Title	GROUNDWATER POTENTIAL MODEL USING STATISTICS AND HYBRID MACHINE LEARNING APPROACHES
By	Mr. Ngoc Thanh Nguyen
Field of Study	Environmental Science
Thesis Advisor	Professor SRILERT CHOTPANTARAT, Ph.D.

Accepted by the GRADUATE SCHOOL, Chulalongkorn University in Partial Fulfillment of the Requirement for the Doctor of Philosophy

.... Dean of the GRADUATE SCHOOL

(Associate Professor YOOTTHANA CHUPPUNNARAT, Ph.D.)

DISSERTATION COMMITTEE

.... Chairman

(Associate Professor Tunlawit Satapanajaru, Ph.D.)

.... Thesis Advisor

(Professor SRILERT CHOTPANTARAT, Ph.D.)

.... Examiner

(Professor MONTRI CHOOWONG, Ph.D.)

.... Examiner

(Assistant Professor PATAMA SINGHRUCK, Ph.D.)

.... Examiner

(Assistant Professor SARAWUT SRITHONGOUTHAI, Ph.D.)

.... Examiner

(SUMET PHANTUWONGRAJ, Ph.D.)

งอค ธานห์ เหวียน :
แบบจำลองศักยภาพน้ำบาดาลโดยวิธีทางสถิติและการเรียนรู้ของเครื่องแบบไฮบริด. (
GROUNDWATER POTENTIAL MODEL USING STATISTICS AND HYBRID
MACHINE LEARNING APPROACHES) อ.ที่ปรึกษาหลัก : ศรีเลิศ โชติพันธรัตน์

การศึกษาวิจัยนี้เป็นการดำเนินการเพื่อวิเคราะห์ศักยภาพของน้ำบาดาลในจังหวัดกาญจนบุรี ประเทศไทย โดยพิจารณาจากปริมาณน้ำใต้ดิน ความเสี่ยงในการปนเปื้อนของน้ำใต้ดิน และคุณภาพน้ำใต้ดิน ในการศึกษานี้ได้สร้างแบบจำลองขึ้นโดยการรวมกระบวนการลำดับชั้นเชิงวิเคราะห์ อัตราส่วนความถี่ และการสุ่มผืนป่า เพื่อประเมินการกระจายเชิงพื้นที่ของทรัพยากรน้ำใต้ดิน นอกจากนี้ วิธีการแบบผสมผสานแบบใหม่ยังได้รับการพัฒนาโดยใช้องค์ประกอบของ maximum entropy และกระบวนการลำดับชั้นเชิงวิเคราะห์เพื่อระบุความเสี่ยงในการปนเปื้อนของนิกเกิลในน้ำใต้ดินอีกด้วย โมเดลการเรียนรู้ในสี่รูปแบบ ซึ่งประกอบด้วย Random Forest - Cross validation, Random Forest – Bootstrap, Artificial Neural Network - Cross validation และ Artificial Neural Network - Bootstrap ถูกนำมาใช้เพื่อประเมินคุณภาพน้ำใต้ดิน ผลการวิจัยพบว่าวิธีการแบบผสมผสานดีกว่าแบบจำลองเดี่ยว และพื้นที่ศักยภาพของน้ำใต้ดินในระดับต่ำและระดับปานกลาง (น้ำใต้ดินในระดับ > 10 m^3/h) กระจายอยู่ทางตะวันตกของกาญจนบุรีเป็นหลัก ในขณะที่ภาคตะวันออกมีศักยภาพน้ำใต้ดินในระดับสูง (m^3/h) ในส่วนของความเสี่ยงในการปนเปื้อนนิกเกิลพบว่าโมเดลแบบผสมผสานระหว่าง maximum entropy และกระบวนการลำดับชั้นเชิงวิเคราะห์ให้ประสิทธิภาพสูงด้วยพื้นที่ใต้เส้นโค้ง 0.86 และความแม่นยำ 0.85 การแสดงผลจากแผนที่ความเสี่ยงการปนเปื้อนของนิกเกิลในน้ำใต้ดินพบว่าประมาณ 24.79% ของกาญจนบุรีตะวันออก (1691.82 ตารางกิโลเมตร)

สาขาวิชา   วิทยาศาสตร์สิ่งแวดล้อม   ลายมือชื่อนิสิต ....
              (สหสาขาวิชา)
ปีการศึกษา   2565   ลายมือชื่อ อ.ที่ปรึกษาหลัก ....

6283008920 : MAJOR ENVIRONMENTAL SCIENCE

KEYWORD:

Ngoc Thanh Nguyen : GROUNDWATER POTENTIAL MODEL USING STATISTICS AND HYBRID MACHINE LEARNING APPROACHES. Advisor: Prof. SRILERT CHOTPANTARAT, Ph.D.

The present study was conducted to delineate the groundwater potential in Kanchanaburi Province, Thailand based on groundwater yield, groundwater contamination risk, and groundwater quality. In this study, an ensemble model was created by combining Analytical Hierarchy Process, Frequency Ratio, and Random Forest to evaluate the spatial distribution map of the groundwater resources. Additionally, a new hybrid approach was developed based on maximum entropy and analytical hierarchy process to delineate the Ni contamination risk in groundwater. Finally, four machine-learning models, including Random Forest - Cross validation, Random Forest – Bootstrap, Artificial Neural Network - Cross validation, and Artificial Neural Network - Bootstrap were used to decipher groundwater quality. The results indicated that the ensemble model was better than individual models in delineating groundwater yield potential. Poor and moderate potential with groundwater yield > 10 m^3/h was distributed mainly in the western Kanchanaburi, while the eastern regions showed high groundwater yield potential (bao nhiêu m^3/h). In terms of contamination risk, the hybrid model between maximum entropy and analytical hierarchy process gave a high performance with an Area Under Curve of 0.86 and Accuracy of 0.85. The map of Ni contamination risk in groundwater showed that approximately 24.79% of the eastern Kanchanaburi (1691.82 km^2) was a very low contamination risk of Ni.

Field of Study: Environmental Science   Student's Signature ....

Academic Year: 2022   Advisor's Signature ....

ACKNOWLEDGEMENTS

Ngoc Thanh Nguyen

TABLE OF CONTENTS

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

Chapter 1. INTRODUCTION

1.1. Statement of the problem

Water is vital for the existence of all life on the Earth and an essential component of several manufacturing operations in a nation. In recent decades, freshwater shortages in many regions have demanded research on the sustainable use of water resources (Chen et al., 2020; Hou et al., 2020; Zehtabiyan-Rezaie et al., 2019), based on the effects of weather elements and human activities on these resources. The intricate process of climate change has caused widespread drought, changes in location, timing form, the number of precipitations, and tropical storm intensity in many areas of the world (Dai, 2013). Rising sea levels have led to salinity intrusion in soil and groundwater in desert and coastal regions (Boonkaewwan et al., 2020; Howard et al., 2010; Shah et al., 2001). Simultaneously, population growth, agricultural expansion, urbanization, and modern industrial activities have created an increasingly large need for freshwater (Francis, 2011; Howard et al., 2010; McDonald et al., 2011; Wilson et al., 2006). To reduce freshwater shortages, the exploitation and sustainable use of water resources have garnered increasing attention, particularly with regard to groundwater. The development of groundwater potential maps is becoming increasingly important in many scientific fields and nations (Chaminé et al., 2015). In particular, groundwater mapping provides suitable locations for drilling groundwater wells for the purpose of water resource management and agricultural activities (Ahmed and Sajjad, 2018; Elbeih, 2015). Simultaneously, a groundwater potential map can reveal the groundwater potential zone of a given region. Additionally, human development and changes in natural processes lead to negative aspects of groundwater quality (Li et al., 2017; Shah, 2005; White et al., 2007). The presence of heavy metals causes considerable concern in various water uses. Consequently, assessing the concentration of toxic pollutants in groundwater is urgently needed in many regions and countries worldwide.

Constantly advancing remote sensing (RS) and geographic information system (GIS) technology has played a vital role in groundwater hydrology in recent decades. GIS is an appropriate tool for solving spatial attribute problems in many environmental science fields (Beck, 2003; Chaudhary and Kumar, 2018; Chowdhury et al., 2009; Edet et al., 1998; Twumasi and Merem, 2006), particularly hydrology and hydrogeology. The information RS provides for the Earth's surface facilitates groundwater potential research in many regions that do not have available data and complex topography. In recent years, several groundwater studies have utilized RS and GIS technology to map groundwater potential zones in their areas of interest. Some scientists (Ghosh et al., 2016; Gnanachandrasamy et al., 2018; Kallali et al., 2007; Machiwal et al., 2011; Pinto et al., 2017; Selvam et al., 2016) have used statistical models such as multi-criteria decision analysis (MCDA), multi-influence factor (MIF), analytical hierarchy process (AHP), frequency ratio (FR) model, logistic regression (LR) model, and evidence belief function (EBF). Other researchers have applied machine learning approaches to analyze, calculate, and map the groundwater potential zone (Alizamir et al., 2018; Kamali Maskooni et al., 2020; Lee et al., 2019; Sahoo et al., 2017; Tan et al., 2020). Machine learning models are often used in groundwater studies, including AB-AD Tree, quadratic discriminant analysis, K-neighbor classification, random forest classifier (RFC), and support vector machine (SVM). The purpose of these models is to delineate areas with large groundwater reserves to exploit and reasonably use groundwater. Although some attempts have been made to develop these models for assessing the groundwater potential for a given region, the accuracy of the models has reportedly ranged from 50% to 92% (Abijith et al., 2020; Kamali Maskooni et al., 2020; Pham et al., 2019; Tien Bui et al., 2019) with regard to the establishment of potential groundwater maps. The improvement of groundwater potential mapping models has a significant role in determining the successful location of groundwater wells. This not only helps establish wells but also guides the sustainable use of groundwater resources.

In this study, Kanchanaburi Province in Thailand is considered because its

groundwater situation faces many problems such as drought and contamination (Srirattana et al., 2021). Moreover, groundwater extraction increases rapidly to meet the demand for domestic and agricultural irrigation (Putthividhya and Pipitsombat, 2018). However, the groundwater potential map in terms of quantity and quality in this province is still limited. Hence, it is necessary to assess groundwater potential for various purposes in Kanchanaburi Province, Thailand. The anticipated output of this study will contribute new approaches in groundwater potential study not only for the study area but also for other regions in the world.

1.2. Research Objective

The general objective of the research is to delineate the groundwater potential zones in terms of quantity and quality in the Kanchanaburi Province, Thailand, where groundwater has been widely used by the inhabitant for their various needs.

1.3. Specific objectives

a) to develop a new ensemble model to efficiently evaluate the spatial distribution map of the groundwater resources.

b) to develop a new hybrid model to delineate the heavy metal contamination risk zone in groundwater.

c) to delineate groundwater quality using the Entropy water quality indexes.

1.4. Research questions

1. How is the spatial distribution of groundwater yield in the study area?

2. Which database and models are suitable for the study area?

3. How is the level of heavy metals contamination risk in groundwater in the study area?

4. Is the hybrid model acceptable for heavy metal risk assessment in groundwater?

5. Which model is acceptable for deciphering groundwater quality?

6. What is the suitability of groundwater potential for different uses?

1.5. Hypothesis

1. The new hybrid approaches (AHP, FR, RF) can effectively assess groundwater potential in a case study of Kanchanaburi Province, Thailand.

2. The new hybrid approaches (Maxent and AHP) can effectively delineate the heavy metal contamination risk zone in groundwater in a case study of Kanchanaburi province, Thailand.

3. The machine learning models can effectively decipher groundwater quality in a case study of Kanchanaburi Province, Thailand.

1.6. Expected outcome

1. A map of spatial distribution of groundwater yield in Kanchanaburi, Thailand using frequency ratio, random forest, and analytic hierarchy process.

2. A map of contamination risk zones in groundwater in Kanchanaburi, Thailand.

3. A map of groundwater quality in Kanchanaburi, Thailand.

Chapter 2. THEORETICAL BACKGROUND AND LITERATURE REVIEWS

2.1. Definition of groundwater potential map

Previous literature reviews indicate that the definition of GWP is a general concept depending on the purpose of each research. Some researchers (Andermann et al., 2012; Fischer et al., 2003; Rodell et al., 2007; Yeh et al., 2006) have focused on groundwater storage, while others (Batte et al., 2008; Janakarajan and Moench, 2006; Nampak et al., 2014) have analyzed the groundwater yield in an area. From a storage perspective, GWP is defined as the total amount of water in aquifers that can be stored for a long time (Kebede, 2013). Considering the yield, GWP is the quantity of groundwater that may be extracted from a groundwater aquifer without surpassing the long-term recharge or the chemical and physical integrity of the basin (Greer, 2008). In addition to the above aspects, groundwater quality is also used to delineate GWP zones for different purposes. Hounsinou (2020) considered the extent of saline intrusion of seawater to set the GWP boundaries. Dhar et al. (2015) defined the GWP zones by overlaying the GWP and groundwater quality indexes. Consequently, the definition of GWP is not a specific concept to use uniformly worldwide. The usage of groundwater storage or yield to define GWP remains controversial because they ignore factors, such as groundwater quality, aquifer properties, sensitivity, contamination, and its intended use. The definition of GWP based on the knowledge of the authors can be stated as follows: "Groundwater potential is the volume of groundwater that can be withdrawn from an aquifer for a particular purpose without affecting the groundwater yield and groundwater quality of an aquifer."

2.2. Number of publications in groundwater potential field

The study area statistics around the world provide hydrologists with a general picture of GWP research. It shows the level of interest in groundwater reserves in different parts of the world. It is the basis for international organizations to select and

implement projects related to the conservation of global groundwater resources. In this section, information regarding the number of research papers on GWP and other parameters is presented. The content includes information about the number of national studies. The information given in Fig. 2.1 was collected from two reliable sources, namely Scopus and Web of Science, with keywords "groundwater potential + year," applied by the "title, abstract, and keywords" search functions. A total of 872 and 707 articles were found in Scopus and Web of Science in this field from 2010 to 2020, respectively. Overall, the number of articles related to GWP map research has remarkably increased in this period (Fig. 2.1), with 87 countries of interest according to authors (Fig. 2.2). Fig. 2.2 also shows the level of interest by hydrologists regarding GWP worldwide over the past decade. India, Iran, and Nigeria had a high number of publications in this research field. Only a small number of relevant studies have been conducted in the rest of the countries (Fig. 2.2). Fig. 2.3 shows the top 10 journals of Scopus and Web of Science systems selected by authors for publication between 2010 and 2020, in which Environmental Earth Sciences and Arabian Journal of Geosciences had the highest publication.
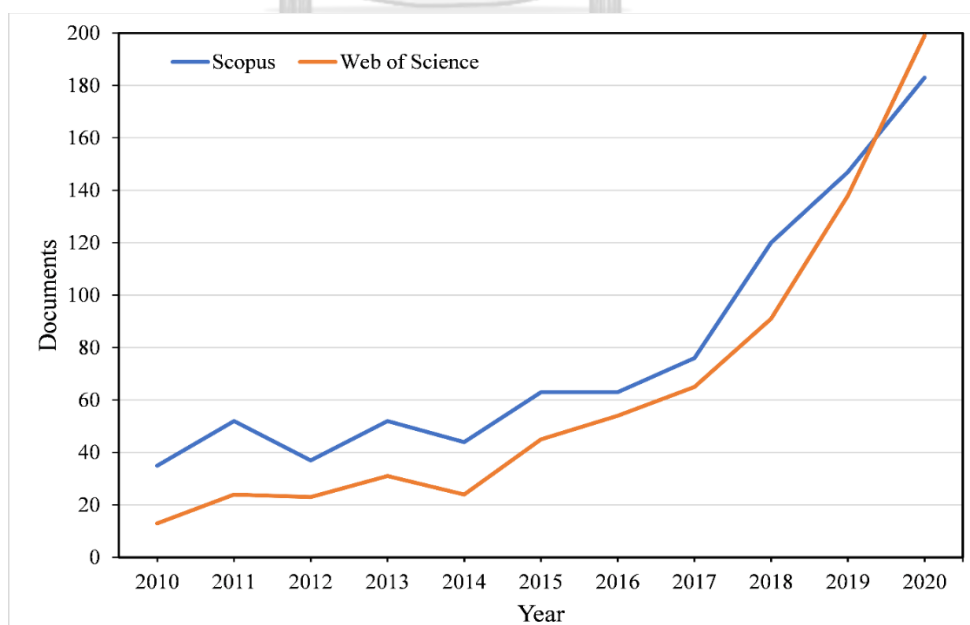
Fig. 2.1. Number of GWP studies in the world between 2010 and 2020
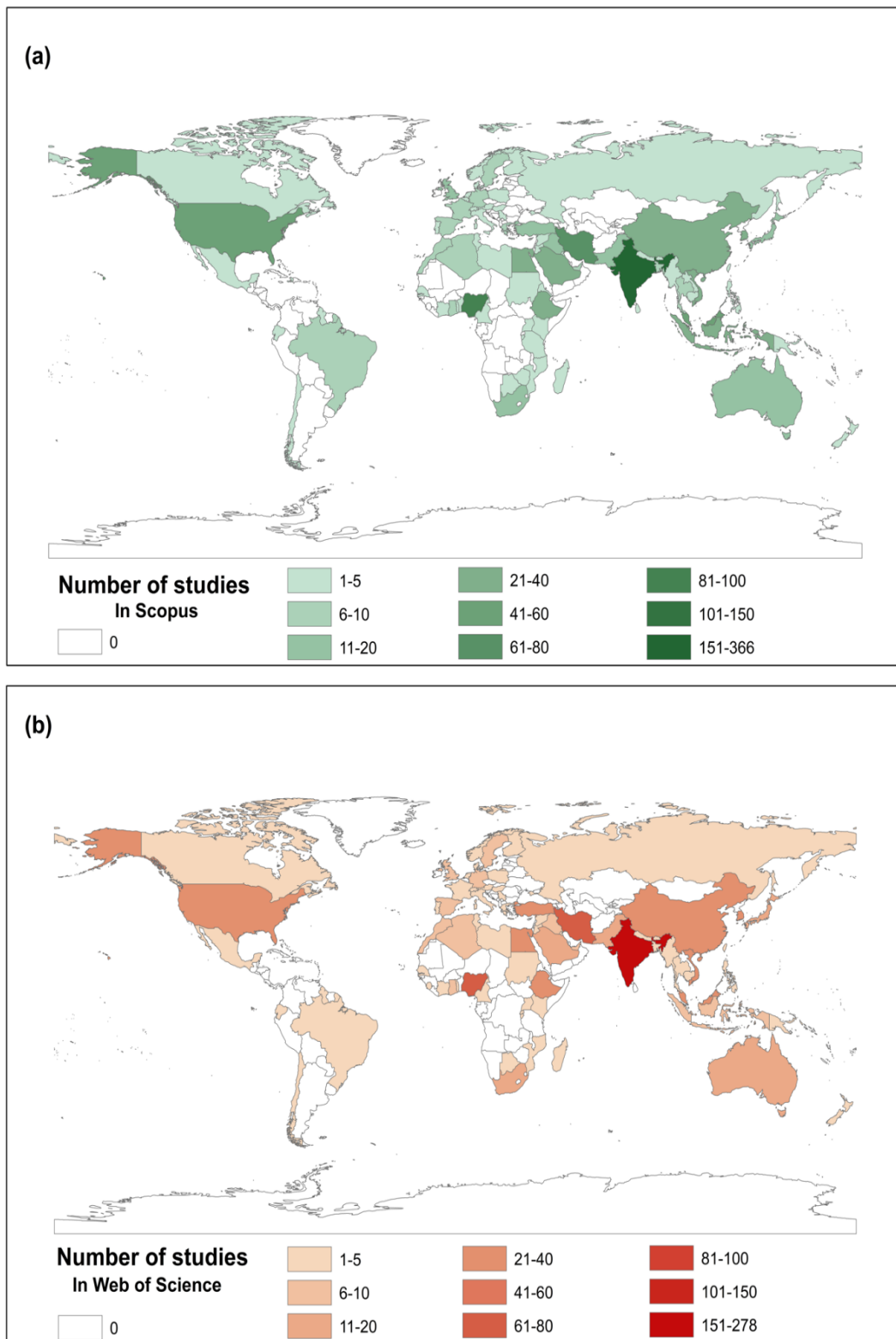
Fig. 2.2. Map of the nations with studies in the world groundwater potential map published from 2010 to 2020 in the Scopus (a) and Web of Science (b) database
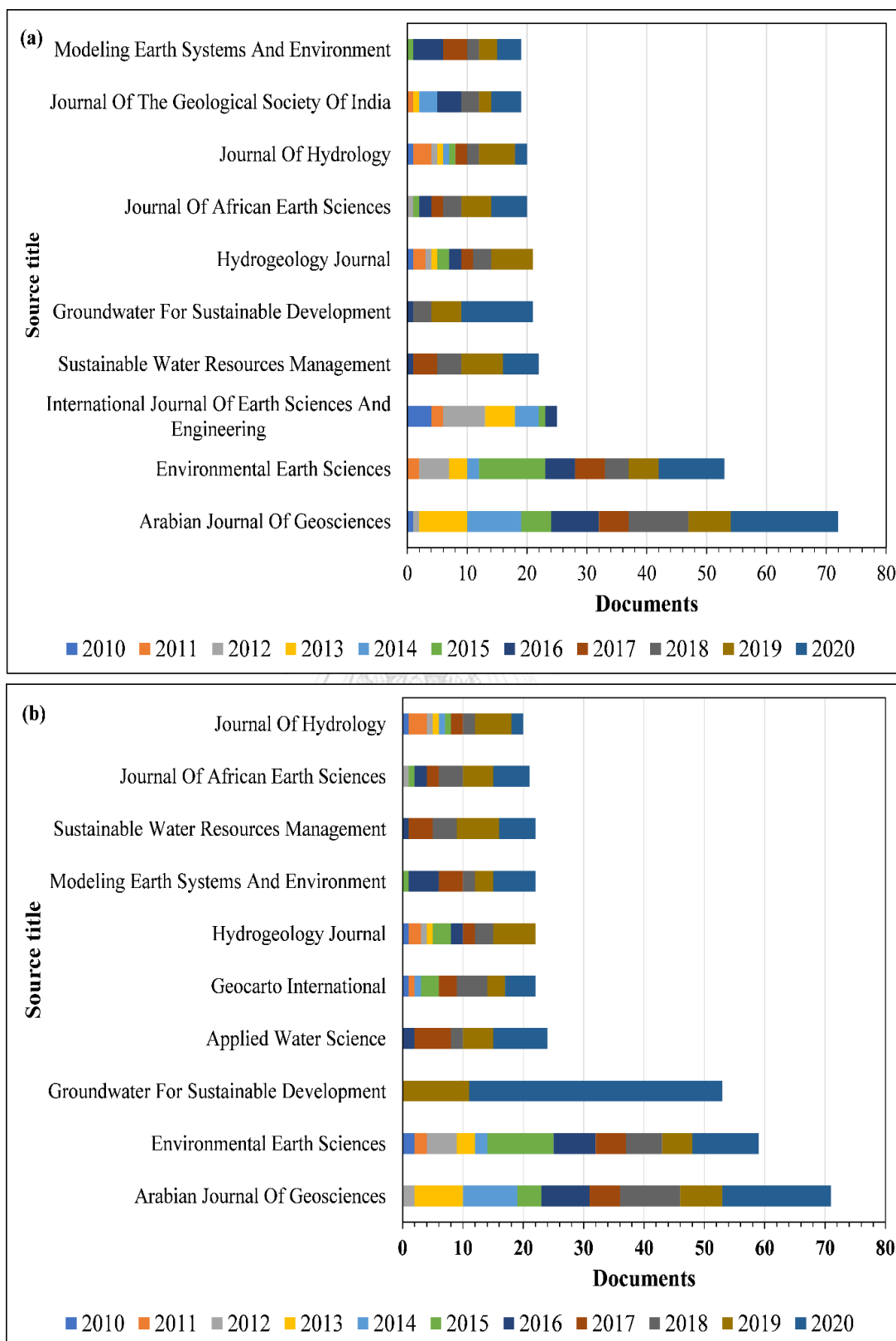
Fig 2.3. Top 10 journals considering number of publications on GWP in the Scopus (a) and Web of Science (b) database from 2010 to 2020

2.3. Parameters, model techniques and validation in groundwater potential studies

The dataset related to this study is compiled from two sources, namely Scopus and Web of Science, and conducted as shown in Fig. 2.4. The first step involves searching the literature. A total of 872 publications on Scopus and 707 publications on Web of Science are related to GWP. The second step is cross-referencing literature between the two systems using DOI and the title of papers. A total of 1085 publications were found in the GWP field from 2010 to 2020. The third step is the calculation of the number of representative samples using Slovin's formula (Marendra and Tangahu, 2020) (Eq. 1).

$$n = \frac{N}{1+Ne^2},$$ (1)

where N is the total of 1085 publications, e is the error margin (0.1), and n is the number of representative samples. The fourth step is setting the dataset. A total of 91 publications are selected for this study based on the citation index from high to low (Table 2.1). The final step is reviewing the content on parameters, models, and validation techniques. This step is the most time-consuming procedure in the research process.
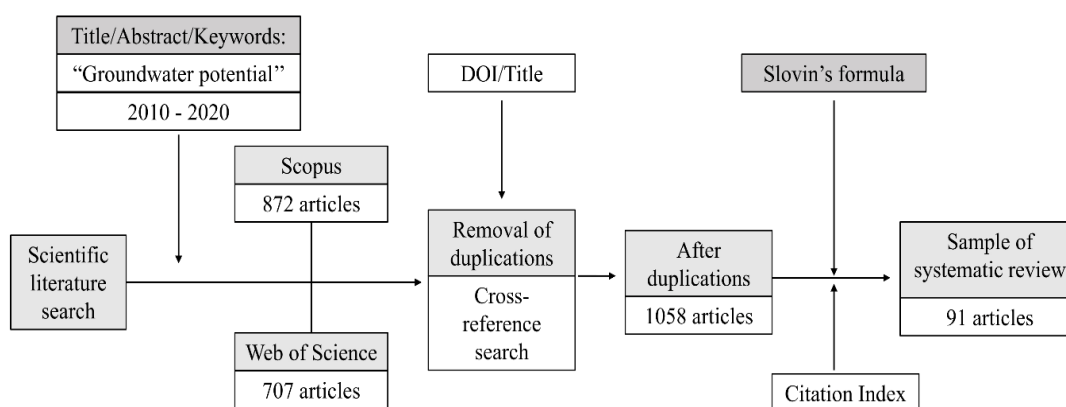
Fig. 2.4. Flow chart of constructing the database

Table 2.1. Database of the reviewed publications

No	Author	Journal	Input parameters	Model (AUC/accuracy)	Validation parameter	Data rate
1	Naghibi et al. (2016)	Environmental Monitoring and Assessment	DS, AS, AL, TWI, LS, PlC, PrC, DrW, DHF, L, LU, DD, LD	BTR (0.810), CART (0.787), RF (0.711),	Non-spring/Spring	70:30
2	Magesh et al. (2012)	Geoscience Frontiers	DS, L, LU, DD, F, D, R, S	MIF (N/A)	N/A	N/A
3	Nampak et al. (2014)	Journal of Hydrology	DS, AL, TWI, C, L, LU, DD, LD, R, S, SPI, NDVI	EBF (0.780), LR (0.72)	GW yield	70:30
4	Oh et al. (2011)	Journal of Hydrology	DS, TWI, C, RD, DrW, L, LD, F, S	FR (0.780)	GW yield	66:34
5	Rahmati et al. (2016)	Catena	DS, AS, AL, TWI, PlC, DrW, L, LU, DD, S	RF (0.831), MIX (0.877)	GW yield	70:30
6	Machiwal et al. (2011)	Water Resources Management	DS, DrW, L, S, Ge, GWD, NR	MCDM-AHP(N/A)	GW yield	NA
7	Rahmati et al. (2015)	Arabian Journal of Geosciences	DS, L, DD, LD, F, R	MCDM-AHP(0.736)	GW yield	N/A
8	Razandi et al. (2015)	Earth Science Informatics	DS, AS, AL, TWI, DrW, L, DD, R, WTL, ATh	FR(0.775), AHP(0.734), CF(0.65)	GW yield	70:30
9	Jha et al. (2010)	Hydrogeology Journal	C, DrW, L, DD, S, Ge, GWD, NR, ATh, AR	MCDM-AHP(0.590)	GW yield	N/A
10	Ozdemir (2011b)	Journal of Hydrology	AS, AL, TWI, LS, C, PlC, PrC, DrF, L, LU, DD, LD, DHD, R, SPI, STI, RP	EBF(0.820)	GW yield	90:10
11	Adiat et al. (2012)	Journal of Hydrology	DS, L, F, D, R	MCDM-AHP(0.812)	GW yield	NA
12	Ozdemir (2011a)	Journal of Hydrology	DS, AS, AL, TWI, C, PlC, PrC, DrF, L, LU, DD, LD, DHD, R, SPI, STI, RP	LR (0.840), FR (0.903), WOE (0.880)	Non-spring/Spring	90:10
13	Naghibi et al. (2017a)	Water Resources Management	DS, AS, AL, TWI, LS, PlC, PrC, RD, DrW, DHF, L, LU, LD, SPI	RF (0.846), GRF(0.856), SVMLL(0.786), SVMP(0.768), SVMIS(0.771), SVMR(0.770)	Non-spring/Spring	70:30
14	Gupta and Srivastava (2010)	Water International	DS, AL, LU, DD, LD	MCDM-AHP(N/A)	N/A	N/A
15	Abd Manap et al. (2014)	Arabian Journal of Geosciences	DS, AL, C, RD, L, LU, F, S	FR(0.747)	GW yield	70:30
16	Naghibi and Pourghasemi (2015)	Water Resources Management	DS, AS, AL, TWI, LS, PlC, PrC, RD, DrW, DHF, L, LU, LD, SPI	BRT(0.861), CART(0.863), RF(0.66), EBF(0.677), GLM(0.772)	Non-spring/Spring	70:30
17	Abd Manap et al. (2013)	Arabian Journal of Geosciences	DS, AL, L, LU, DD, F, R, S, Ge	MCDM-AHP(0.760)	GW yield	N/A
18	Shen et al. (2013)	Canadian Geotechnical Journal	AT,	Correlation Analysis	N/A	N/A
19	Chen et al. (2018)	Science of The Total Environment	DS, AS, AL, TWI, PlC, PrC, DrW, L, LU, S, SPI, STI, NDVI	LR(0.753), WOE(0.700), DT(0.777)	Non-spring/Spring	70:30
20	Naghibi et al. (2017b)	Journal of Hydrology	DS, AS, AL, TWI, PlC, RD, DrW, L, LU, SPI, NDVI	AB(0.926), BIG(0.935), GAM(0.92), NIB(0.844), MARS(0.948)	GW yield	70:30
21	Golkarian et al. (2018)	Environmental Monitoring and Assessment	DS, AS, AL, TWI, LS, PlC, PrC, RD, DrW, DHF, L, LU, LD	RF(0.797), DT(0.773), MARS(0.842)	Non-spring/Spring	70:30
22	Zabihi et al. (2016)	Environmental Earth Sciences	DS, AS, AL, TWI, LS, PlC, PrC, DrW, DHF, L, LU, DD, LD	RF(0.709), MARS (0.732)	Non-spring/Spring	70:30
23	Ghorbani Nejad et al. (2017)	Geocarto International	DS, AS, AL, TWI, C, DrW, DHF, L, LU, DD, LD	EBF(0.762), WOE(0.736)	GW yield	70:30
24	Oikonomidis et al. (2015)	Journal of Hydrology	DS, AS, L, DD, LD, R, GWD, NR, AT, GWQ	MCDM-AHP(N/A)	Quality	N/A
25	Dar et al. (2010)	Journal of Hydrology	C, L, LU, F, S, Ge	MCDM-AHP(N/A)	N/A	N/A

No	Author	Journal	Input parameters	Model (AUC/accuracy)	Validation parameter	Data rate
26	Tahmassebipoor et al. (2016)	Arabian Journal of Geosciences	DS, AL, TWI, C, DtF, L, LU, DD, LD, R, S	EBF(0.837), WOE(0.782)	GW yield	70:30
27	Pinto et al. (2017)	Applied Water Science	DS, AL, L, LU, DD, LD, R, S	MCDM-AHP(N/A)	GW yield	N/A
28	Shekhar and Pandey (2015)	Geocarto International	DS, L, DD, LD, R, S, Ge, AtTh	MCDM-AHP(N/A)	GW yield	N/A
29	Gayo et al. (2012)	Earth-Science Reviews	L, S, NDVI, AT, GWQ	Correlation Analysis	N/A	N/A
30	Agarwal and Garg (2016)	Water Resources Management	DS, L, LU, DD, R, S, Ge, WTL	MCDM-AHP(N/A)	GW yield	N/A
31	Chen et al. (2019a)	Journal of Hydrology	DS, AS, AL, TWI, C, PlC, PrC, DtF, L, LU, R, S, SPI, STI, DtR	ANFIS+TLBO(0.905), ANFIS+BBO(0.887)	Non-spring/Spring	70:30
32	Mukherjee et al. (2012)	Water Resources Management	DS, AL, L, LU, DD, LD, R, S, Ge	MCDM-AHP(N/A)	GW yield	N/A
33	Rahmati et al. (2018)	Journal of Hydrology	DS, AS, AL, TWI, RD, DtW, DtF, L, Cl, RSP, TRI, TPI	LR(0.904), DT(0.868), SVM(0.889)	Non-spring/Spring	70:30
34	Agarwal et al. (2013)	Journal of Earth System Science	DS, L, LU, DD, F, R, S, Ge, GWD	MCDM-AHP(N/A), ANP(N/A)	GW yield	N/A
35	Singh et al. (2011)	Environmental Earth Sciences	DS, L, F, S, Ge, WTL	MCDM-AHP(N/A)	GW yield	N/A
36	Pourghasemi and Beheshtirad (2015)	Geocarto International	DS, AS, AL, TWI, LS, PlC, DtW, DtF, L, LU, DD, LD	EBF(0.817)	Non-spring/Spring	70:30
37	Elewa and Qaddah (2011)	Hydrogeology Journal	DS, L, DD, LD, R, Ge, NR, GWQ	MCDM-AHP(N/A)	GW yield	N/A
38	Arulbalaji et al. (2019)	Scientific Reports	DS, TWI, C, L, LU, DD, LD, R, S, Ge, TRI, TPI	MCDM-AHP(N/A)	GW yield	N/A
39	Chen et al. (2019b)	Natural Resources Research	DS, AS, AL, TWI, PlC, PrC, DtW, L, LU, S, SPI, STI, NDVI, DtR	FLDF(0.845), FLDF+RTF(0.882), FLDF+BG(0.892)	Non-spring/Spring	70:30
40	Miraki et al. (2019)	Water Resources Management	DS, AS, AL, TWI, C, RD, L, LU, LD, R, SPI, NDVI	LR(0.825), NB(0.800), RFC+BTR(0.878)	Non-spring/Spring	70:30
41	Kordestani et al. (2019)	Hydrogeology Journal	DS, AS, AL, TWI, LS, PlC, PrC, RD, DtW, L, LU, SPI	LR(0.755), EBF+BTR(0.821)	Non-spring/Spring	70:30
42	Naghibi et al. (2018)	Theoretical And Applied Climatology	DS, AS, AL, TWI, LS, PlC, PrC, L, LU, SPI	BTR(0.846), RF(0.846), MARS(0.76), SVM(0.83)	Qanat	70:30
43	Ibrahim-Bathis and Ahmed (2016)	Egyptian Journal of Remote Sensing and Space Sciences	DS, LU, DD, LD, R, S, Ge, AT	MCDM-AHP(N/A)	N/A	N/A
44	Abdalla (2012)	Journal of African Earth Sciences	DS, AL, RD, L, F, GWQ	MCDM-AHP(N/A)	N/A	N/A
45	Nag and Ghosh (2013)	Environmental Earth Sciences	DS, LD, Ge	MCDM-AHP(N/A)	N/A	N/A
46	Patra et al. (2018)	Journal of Cleaner Production	DS, AL, L, LU, DD, R, S, NDVI, Ge, GWD, NR	MCDM-AHP(0.750)	Groundwater level	N/A
47	Arabameri et al. (2019)	Science of The Total Environment	DS, AS, AL, TWI, C, DtW, DtF, L, LU, DD, LD, R, S, SD, DtR	MIF(0.873), LR(0.905), AHP(0.87), WOE(0.885)	GW yield	70:30
48	Thapa et al. (2017)	Applied Water Science	L, LD, Ge	MIF(0.780)	GW yield	N/A
49	Mogaji et al. (2015)	Arabian Journal of Geosciences	DS, L, DD, LD, F, R, S	MCDM-AHP(0.857)	GW yield	N/A
50	Gumma and Pavelic (2013)	Environmental Monitoring and Assessment	DS, L, LU, DD, R, S, Ge	MCDM-AHP(N/A)	Groundwater level	N/A

No	Author	Journal	Input parameters	Model (AUC/accuracy)	Validation parameter	Data rate
51	Fenta et al. (2015)	Hydrogeology Journal	DS, L, LU, DD, LD, R, Ge	MCDM-AHP(0.740)	GW yield	N/A
52	Kumar et al. (2014)	Water Resources Management	DS, L, LU, DD, LD, R, S	MCDM-AHP(N/A)	Groundwater depth	N/A
53	Khosravi et al. (2018)	Hydrology and Earth System Sciences	DS, AS, AL, TWI, PlC, DtW, DtF, L, LU, R, S, SPI, TRI	ANFIS+WD (0.873), ANFIS+DE(0.875), ANFIS+FA(0.873), ANFIS+PSO(0.865), ANFIS+BA(0.839)	N/A	70:30
54	Mallick et al. (2015)	Hydrological Processes	AL, DtW, L, LU, DD, F, S, Ge, GWD, ST	MCDM-AHP(N/A)	Discharge/groundwater depth	N/A
55	Preeja et al. (2011)	Journal of The Indian Society of Remote Sensing	DS, L, LU, DD, LD, S, Ge	MCDM-AHP(N/A)	N/A	N/A
56	Lee et al. (2018)	Geocarto International	DS, AL, RD, DtW, DtF, L, LU, LD, S, SPI, GWD	SVM(0.808)	GW yield	50:50
57	Singh et al. (2018)	Ecological Indicators	DS, DtW, L, LU, DD, S, RC	MCDM-AHP(0.820)	GW yield	N/A
58	Rajaveni et al. (2017)	Applied Water Science	DS, L, LU, DD, LD, Ge	MCDM-AHP(N/A)	Groundwater level	N/A
59	Naghibi and Moradi Dashtpagerdi (2017)	Hydrogeology Journal	DS, AS, AL, TWI, LS, PlC, PrC, RD, DtW, DtF, L, LU, LD, SPI	MARS(0.814), LDA(0.792), QDA(0.805), KNN(0.790)	Non-spring/Spring	70:30
60	Rahmati and Melesse (2016)	Science of The Total Environment	DS, AL, TWI, RD, DtW, L, LU, LD, S, GWQ	Dempster-Shafer(0.877)	GW yield	70:30
61	Dar et al. (2011)	Environmental Monitoring and Assessment	L, LU, DD, LD, Ge, RL	MCDM-AHP(N/A)	Non-spring/Spring	N/A
62	Pham et al. (2019)	Catena	DS, AS, AL, TWI, PlC, RD, L, LU, DtD, S	RF+DS(0.922),MB+DS(0.812),BG+DS(0.828)	GW yield	70:30
63	Andermann et al. (2012)	Journal of Hydrology-Regional Studies	DS, L, LU, LD, F, S, Ge	MCDM-AHP(0.705)	GW yield	N/A
64	Jenifer and Jha (2017)	Journal of Hydrology	DS, AL, DtW, L, DD, LD, R, S	MX(0.51), AHP(0.87), Catastrophe theory(0.460)	Groundwater level	N/A
65	Worqlul et al. (2017)	Applied Geography	DS, LU, R, S, GWD, DtR, FD	MCDM-AHP(N/A)	N/A	N/A
66	Selvam et al. (2015)	Environmental Earth Sciences	DS, L, LU, DD, LD, R, S	MCDM-AHP(N/A)	N/A	N/A
67	Deepika et al. (2013)	Environmental Earth Sciences	DS, L, F, DtD, S, Ge, RL	MCDM-AHP(N/A)	N/A	N/A
68	Avinash et al. (2011)	Geocarto International	RD, L, Ge, TRI, RL	Correlation Analysis	N/A	N/A
69	Falah et al. (2017)	Geocarto International	DS, AS, AL, TWI, PlC, DtW, L, LU, DD, LD, S	FR(0.837), WOE(0.763), GAM(0.777), EBFTM(0.854)	Non-spring/Spring	70:30
70	Kumar and Krishna (2018)	Geocarto International	DS, L, DD, LD, R, S, Ge, ATh	MCDM-AHP(0.754)	GW yield	N/A
71	Gaur et al. (2011)	International Journal of Applied Earth Observation and Geoinformation	DS, L, LU, D, S	MCDM-AHP(N/A)	GW yield	N/A
72	Das (2019)	Groundwater for Sustainable Development	DS, L, LU, DD, LD, R, S, SD, Ge	MIF(0.710), EBF(0.750), AHP(0.700)	Non-spring/Spring	N/A
73	Ahmed et al. (2015)	Earth Science Informatics	DS, L, DD, R, S	Catastrophe theory	GW yield	N/A
74	Mudingami et al. (2012)	Physics and Chemistry of The Earth	AR,	Correlation Analysis	GW yield	N/A
75	Murmu et al. (2019)	Groundwater for Sustainable Development	DS, L, LU, DD, LD, R, S, Ge	MCDM-AHP(N/A)	GW yield	N/A

No	Author	Journal	Input parameters	Model (AUC/accuracy)	Validation parameter	Data rate
76	Guru et al. (2017)	Journal of King Saud University Science	DS, L, LU, DD, LD, Ge, WTL	FR(0.772)	Non-spring/Spring	70:30
77	Mousavi et al. (2017)	Aims Geosciences	DS, AS, AL, TWI, PlC, PrC, RD, DtW, DtF, L, LU, LD	BRT(0.872), FR(0.832)	Non-spring/Spring	70:30
78	Ghosh et al. (2016)	Modeling Earth Systems and Environment	DS, L, LU, F, D, S	MCDM-AHP(N/A)	GW yield	N/A
79	Singh et al. (2013)	Hydrological Sciences Journal	DS, L, DD, F, Ge	MCDM-AHP(N/A)	GW yield	N/A
80	Mandal et al. (2016)	Water Resources Management	DS, DtW, L, LU, DD, LD, R, S, NDVI, Ge, NR, AT	MCDM-AHP(N/A)	GW yield	N/A
81	Aizebeokhai et al. (2010)	Environmental Earth Sciences	AR	Correlation Analysis	GW yield	N/A
82	Naghibi et al. (2019)	Environmental Monitoring and Assessment	DS, AS, AL, TWI, C, PlC, PrC, RD, DtW, DtF, L, LU, LD	BRT(0.898), CART(0.869), RF(0.901), EBFTM(0.904), RTF(0.862)	Non-spring/Spring	70:30
83	Saha (2017)	Spatial Information Research	DS, TWI, RD, DtW, L, LU, DD, LD, R, S, SPI, RL	MCDM-AHP(0.777)	Pumping wells/ Groundwater depth	N/A
84	Nguyen et al. (2020b)	Applied Sciences-Basel	AS, AL, TWI, LS, C, RD, L, LU, R, S, STI	LR(0.710), LR+DLR(0.770), LR+BG(0.735), LR+RS(0.744),LR+CG(0.715)	GW yield	70:30
85	Sameen et al. (2019)	Natural Resources Research	DS, AS, AL, TWI, C, PlC, PrC, DtW, L, LU, SPI, STI, TRI	RF(0.500), AB(0.580), SVM (0.500), PDA(0.460), SRF (0.83), ANN+RF(0.45)	GW yield	70:30
86	Kumar et al. (2016)	Sustainable Water Resources Management	DS, L, LU, S, Ge	MCDM-AHP(N/A)	GW yield	N/A
87	Al-Abadi (2015)	Environmental Earth Sciences	DS, AL, DtF, L, LU, GWD, AT, DtR	WOE(0.800)	GW yield	70:30
88	Termeh et al. (2019)	Hydrogeology Journal	DS, AS, AL, TWI, LS, PlC, PrC, DtW, DtF, L, LU, DD, LD, R, S	EBP(0.800), FR(0.860), MX(0.862), ANFIS+GA(0.915), ANFIS+BBO(0903), ANFIS+SA(0.83), ANFIS(0.82)	GW yield	70:30
89	Mohan et al. (2018)	Hydrology and Earth System Sciences	LU, R	MCDM-AHP(N/A)	GW yield	N/A
90	Elmahdy and Mohamed (2014)	Geocarto International	DS, DtF, DtD, Cl, RL	MCDM-AHP(N/A)	GW yield	N/A
91	Bagyaraj et al. (2013)	Frontiers of Earth Science	DS, L, DD, LD, Ge	MCDM-AHP(N/A)	N/A	N/A

Abbreviations of input parameters: DS – Slope (degree/percentage); AS – Slope aspect; AL - Altitude/Elevation; TWI - Topographic wetness index; LS – Slope length; C – Curvature; PlC – Plan curvature; PrC - Profile curvature; RD - River density; DtW - Distance to water bodies/rivers; DtF – Distance to faults; L – Geology (Lithology); LU - Land use; DD - Drainage density; LD - Lineament density/Fault density; F - Fault/Lineament; D - Drainage; DtD - Distance to drainage; R - Rainfall/Precipitation; S – Soil; SD – Soil depth; SPI - Stream power index; STI - Sediment transport index; NDVI - Normalized difference vegetation index; Ge - Geomorphology; GWD - Groundwater depth; NR - Net recharge; WTL - Water table level; ATh - Aquifer thickness; AR - Aquifer resistivity; RP - Relative permeability map; AT - Aquifer type; DtR - Distance to roads; GWQ - Groundwater quality; CI - Convergence index; RSP - Relative slope position; TRI - Terrain ruggedness index; TPI - Topographic position index; ST - Surface temperature; RC - Runoff coefficient; RL - Relief; FD - Flow direction. **Abbreviations of models:** AB - AdaBoost; AHP - Analytic hierarchy process; ANFIS - Adaptive neuro-fuzzy inference system; ANN - Artificial neural network; ANP - Analytical network process; BA - Bees algorithm; BBO - Biogeography-based optimization; BG – Bagging; BTR - Boosted regression tree; CART - Classification and regression trees; CF - Certainty factor; CG - Cascade generalization; DE - Differential evolution; DLR – Dagging; DS - Decision Stump; DT - Decision Tree; EBF - Evidential belief function; EBFTM - Evidential belief function and tree-based; FA - Firefly algorithm; FDA - Flexible discriminant analysis; FLDF - Fisher's linear discriminant function; FR - Frequency-ratio; FREM - Frequency ratio data mining ensemble model; GAM - Generalized additive model, GLM - General linear model; GRF - Genetic algorithm optimized random forest; IWO - Invasive weed optimization; KNN - K-nearest neighbor; LDA - Linear discriminant analysis; LR - Logistic regression; MARS - Multivariate adaptive regression spline; MB - MultiBoost; MCDM - Multi-criteria decision making; MIF - Multi influencing factor; MX - Maximum entropy; NB - Naïve Bayes; PDA - Penalized discriminant analysis; PSO - Particle swarm optimization; QDA - Quadratic discriminant analysis; RF - Random forest; RS - Random subspace; RSE - Random Subspace Ensemble; RTF - Rotation forest; SI - Statistical Index; SRF - Self-Learning random forest; SVM - Support vector machine; SVML – SVM – linear; SVMP – SVM- polynomial; SVMR - SVM – radial; SVMS - SVM – sigmoid; TLBO – Teaching learning based optimization; WOE - Weights of evidence.

2.3.1. Input parameters in groundwater potential studies

GWP studies are generally developed on the basis of climate, geology, hydrology, land cover, topography, and aquifer-related data. A total of 41 input parameters are related to the identification of groundwater-rich areas (Fig. 2.5). The number of input parameters necessary to establish the GWP map is different in each study. This number ranges from 2 to 17 parameters. Input parameters are the required information of a GWP model. These parameters provide databases related to GWP and are the basis for model application. The frequency of input parameters used in GWP studies is depicted in Fig. 2.5. These data are usually extracted from conventional data, existing maps, remote sensing, and survey. Eight common factors, including geology, slope, land use, soil type, drainage density (DD), fault/lineament density, altitude, and rainfall with a usage frequency larger than 50%, are available in GWP studies during 2010–2020.
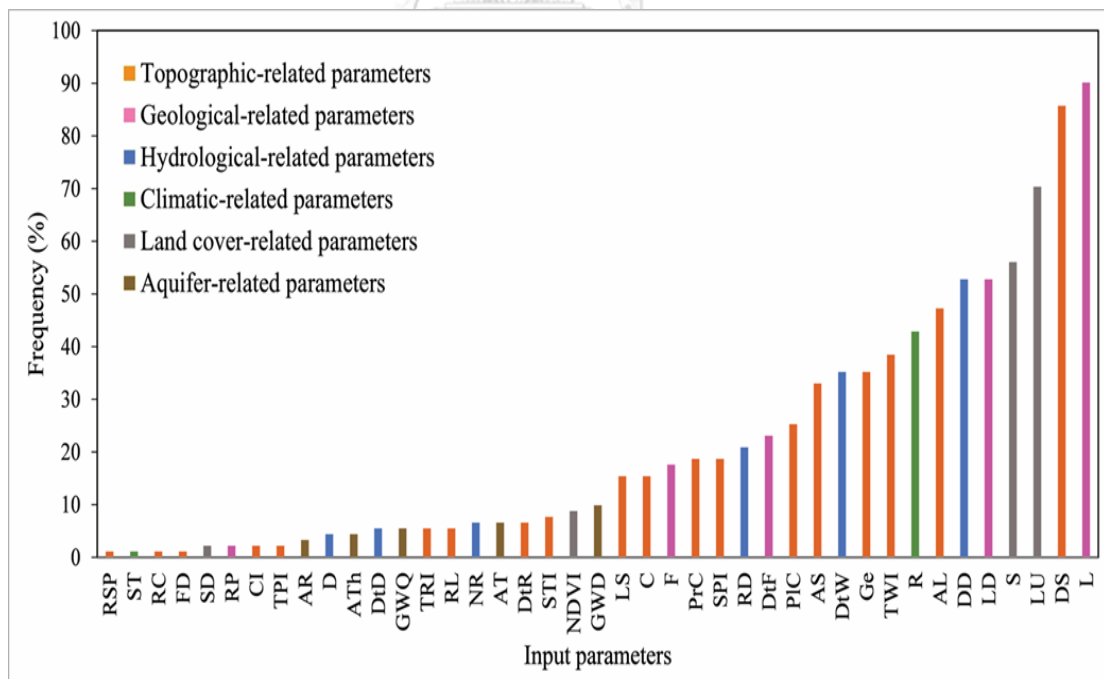
Fig. 2.5. Frequency of factors used in GWP studies from 2010 to 2020

2.3.1.1. Topographic-related parameters

Topographic-related parameters are factors that describe the topographical features of a study area. The frequently used topographic parameters in GWP research include the following: slope, slope aspect, altitude, topographic wetness index (TWI), slope length, curvature, plan curvature, profile curvature, stream power index, sediment transport index, geomorphology, distance to roads, convergence index, relative slope position, terrain ruggedness index, flow direction, and relief. Furthermore, five topographic-related parameters, namely slope, slope aspect, altitude, TWI, and slope length, were widely utilized in most GWP studies (Fig. 2.5). Slope controls the percolation of water into the soil (Ali et al., 2015). Magesh et al. (2012) proved that a gentle slope means that surface water has less residence time on the ground. This leads to water having more time to seep into the ground, whereas a high slope increases flow speed and erosion, and therefore, rainwater does not have sufficient time to infiltrate. Areas with low slopes have negative surface runoff and positive percolation rates, whereas areas with high slopes usually have high amounts of soil runoff and fast meteoric water evacuation by direct hitting water/or rainfall. Adiat et al. (2012) concluded that groundwater recharge depends on the surface flow velocity and vertical percolation controlled by slope degree. Naghibi and Pourghasemi (2015) indicated that the slope aspect also affects hydrological processes because it determines rainfall direction, moisture, plant growth, and snow melting. Tahmassebipoor et al. (2016) determined that groundwater is often limited at high altitudes and abundant at low elevations. Rahmati et al. (2016) applied TWI to measure the amount of runoff accumulation at any place in a basin. GWP partly explains a strong inverse correlation between the TWI index and groundwater yield. Naghibi et al. (2017a) indicated that the possibility of soil loss also impacts GWP, which is calculated on the basis of slope length. GWP is also high when the slop length value is low. Overall, topographic-related parameters control permeability, surface flow direction, and number of precipitations

seeping into the soil. In the last decade, the growth of the aerospace industry has greatly contributed to many scientific fields. Digital elevation models (DEMs) are useful for determining the topography of a certain area (Agarwal and Garg, 2016; Pradhan et al., 2020; Rajasekhar et al., 2019). DEM is usually taken from the Shuttle Radar Topography Mission (SRTM) (Arkoprovo et al., 2012; Nanda et al., 2017; Prasad et al., 2020) or ASTER sensors (Mosavi et al., 2020; Sachdeva and Kumar, 2020; Waikar and Nilawar, 2014) and stored in a raster format. Topographic data are normally extracted from digital elevation information.

2.3.1.2. Geological-related parameters

Geological-related parameters are commonly accepted to govern the potential of groundwater. In GWP mapping studies, distance to faults, geology, lineaments, lineament density, and relative permeability factors are classified among the geological group (Ozdemir, 2011a). Among these factors, geology is the first aspect chosen by researchers in building a GWP model, followed by lineament density (Fig. 2.5). This selection is understandable because differences in lithologies, structure, folds, and faults result in changes in the strength and permeability of soil and rock (Ozdemir, 2011a). Geology plays a significant role in finding water underground. This factor has been mentioned in many groundwater studies. From a geological perspective, groundwater exists in the fractures and voids of consolidated rock as well as in the porosity of unconsolidated sediment. Therefore, GWP research considers the water retention condition of the geological features, such as weathered products, source rocks, and unconsolidated or consolidated sediments (Xie et al., 2014). According to Termeh et al. (2019), hydraulic conductivity, porosity, and groundwater flow of an aquifer are dominated by lithology features. Assatse et al. (2016) pointed out that unconsolidated sediment rocks usually have large gaps, facilitating the existence and movement of water. Areas of unconsolidated sediment

rock have high GWP. In addition, some consolidated sediment rocks are also deemed beneficial for groundwater storage. Consolidated sediment rocks are normally sandstone and limestone (Assatse et al., 2016). In contrast, some rock types, such as metamorphic rock and volcanic rocks, are not generally considered adequate materials for good GWP. Thus, geological structure determines the existence of groundwater in a given region. To collect geological data, most groundwater studies have utilized field surveys (Bagyaraj et al., 2019) or are based on a geologic map (Zabihi et al., 2016). In addition to geological features, lineaments are also a preferential option in GWP studies. Lineaments, which are surficial expressions of faults, are an expression of underlying geological structures. Additionally, lineaments occur in the forms of fractures, lithological limits, and fault zones on the ground. Lineaments often appear in mountainous areas or straight coastlines (Adiat et al., 2012; Golla, 2020; Martha et al., 2013; Rahmati et al., 2015). In the past, lineaments were usually extracted from geology and topography maps (Moore and Waltz, 1983; O'leary et al., 1976). However, more recently, the development of space science and aerial data (such as radar, Landsat ETM (Enhanced Thematic Mapper), and Aster DEM (Digital Elevation Model)) has allowed the collection of lineaments from satellite photographs (Das et al., 2018; Hashim et al., 2013; Mwaniki et al., 2015). (Akinluyi et al., 2018) demonstrated that GWP is influenced by lineaments. Elements belonging to lineaments such as foliation, fractures, and faults are commonly used to find suitable groundwater exploitation locations. The factors related to lineament, such as distance to faults and lineament density, represent the relationship between surface and subsurface water flow through water mobility and infiltration (Termeh et al., 2019). Abdalla (2012) argued that the occurrence of faults is a good condition for groundwater recharge, especially in hard rock locations. Indeed, the fractures, joints, and lineaments act as a conduit, which strengthens the connection between surface water and groundwater. A fault zone can act as a conduit for vertical subsurface flow

or as a barrier for horizontal flow, thus making GWP poor (Bense and Person, 2006). Lineament density reflects the groundwater phenomenon under the Earth's surface. High lineament density corresponds to high secondary porosity (Haridas et al., 1998); hence, aquifers and aquitards are quickly recharged through water infiltration. Concurrently, groundwater movement and retention may be calculated via lineament density. As such, lineament data are used as an indispensable part of finding water potential underground. The first lineament data were explored in a groundwater study carried out by (Lattman and Parizek, 1964). Subsequently, many researchers have applied this approach to complex terrain areas (Abijith et al., 2020; Al-Ruzouq et al., 2019; Ali et al., 2015; Ibrahim-Bathis and Ahmed, 2016; Kamali Maskooni et al., 2020).

2.3.1.3. Hydrological-related parameters

Hydrological-related parameters play a vital role in delineating GWP zones of a given region. River density, distance to rivers, drainage, DD, distance to drainage, and net recharge are hydrological-related parameters in GWP. Among these parameters, DD and distance to rivers were more commonly used than other hydrological factors (Fig. 2.5). Rahmati et al. (2016) reported that GWP, groundwater recharge, and hydrogeological systems are affected by hydrological parameters. High groundwater productivity is rarely found in low river density areas that are far from rivers, streams, and surface water. Chen et al. (2018) indicated that surface water sources are the primary recharging sources, which affect the formation capability of groundwater springs. In addition to river-related factors, drainage-related factors also affect the GWP of an area. DD represents the drainage capacity of a basin based on the length of water flow (Bagyaraj et al., 2013; Martínez-Santos and Renard, 2020; Singh et al., 2013). DD represents residence times of groundwater (Das, 2019), where a high DD implies a large water loss capability and vice versa (Arkoprovo et al., 2012). DD is expressed by the presence of flows on the earth's surface, such as rivers and

streams. The numbers of rivers are often distributed more in delta regions than in hill regions (Mukherjee et al., 2012). DD has been directly designated as GWP; the greater this value, the higher the GWP (Nasir et al., 2018). According to Strahler (1964), DD is calculated by the overall length of rivers and outflows in a given locality divided by the total surface area of the locality. According to Moglen et al. (1998), source rock and climate features determine the shape of stream systems. Areas with a hydrology system with short and scattered rivers have a high GWP and vice versa. Ghosh et al. (2016) observed the inverse correlation of DD and permeability. The zones of high DD are associated with decreased permeability and significant surface runoff. Therefore, the GWP in these zones is low. The length factor of river and flow systems directly affects the DD of a region. In addition to the length of the flow system, other factors also impact DD, including lithology, runoff, vegetation cover, and infiltration. Therefore, DD is a necessary element in groundwater studies (Konkul et al., 2014). It is regarded as an index of groundwater recharge (Gupta and Srivastava, 2010; Mandal et al., 2016; Mosavi et al., 2020; Thapa et al., 2018). Similar to the slope, DD is usually extracted from DEM data (typically ASTER GDE, and SRTM 1 Arc Second Global satellite images with a spatial resolution of 30 m using ArcGIS 10.5 software) (Bagyaraj et al., 2013; Ibrahim-Bathis and Ahmed, 2016; Waikar and Nilawar, 2014).

2.3.1.4. Climatic-related parameters

Climatic-related parameters play a significant role in groundwater formation. In the last decade, GWP models have considered precipitation and surface temperature as determining factors in the GWP in a study area (Elewa and Qaddah, 2011; Mallick et al., 2015; Razandi et al., 2015; Shekhar and Pandey, 2015). Rainfall is a significant factor affecting groundwater recharge (Mukherjee et al., 2012; Owor et al., 2009; Shekhar and Pandey, 2015) and its occurrence (Gumma and Pavelic, 2013; Mosavi et al., 2020; Mukherjee et al., 2012; Nguyen et al., 2020b; Pham et al., 2019) in a study area. In some hydrological studies (Abijith et al., 2020; Klongvessa et al., 2018; Martínez-Santos

and Renard, 2020; Rahmati et al., 2015; Suganthi et al., 2013; Thapa et al., 2018), rainfall is also a factor that influences the recharge capacity of an area. Precipitation contributes a large amount of water to aquifers and aquitards through subsurface infiltration systems. As the rainfall increases, the groundwater level also rises (Chotpantarat et al., 2014; Shekhar and Pandey, 2015). The groundwater recharge potential is generally higher in the rainy season than in the dry season, thus increasing groundwater level (Konkul et al., 2014; Wisittammasri and Chotpantarat, 2016). This finding shows that precipitation is the water supply source for groundwater (Agarwal and Garg, 2016). In addition to the precipitation factor, Mallick et al. (2015) also utilized the surface temperature factor in the GWP study. They assumed that the heat signatures of the earth's surface help discover GWP based on specific heat. For example, dry soil has a lower heat capacity than saturated soil. The surface temperature was usually calculated from satellite imagery (Mallick et al., 2015).

2.3.1.5. Land cover-related parameters

Land use, soil type, soil depth, and NDVI are assigned to the land cover-related group in the present study. Fig. 2.5 reveals that land use and soil type are the two commonly used land cover factors by hydrologists in the past decade. Land use describes the various land-use units of humans on the earth's surface, which presents the domination of recharge rate and groundwater usage (Chen et al., 2018). Therefore, land use plays a major role in groundwater exploitation and usage (Bagyaraj et al., 2019; Dar et al., 2010; Mandal et al., 2016; Mukherjee et al., 2012; Waikar and Nilawar, 2014). Zones affected by urbanization and agricultural activities often require the exploitation of large amounts of groundwater (Odeh et al., 2019). Other zones such as forests and water bodies are rarely intervened by humans (Lone et al., 2013). Consequently, human interference directly affects groundwater storage and recharge capacity. Chen et al. (2018) reported that each land use type has a difference in water permeability. Built-up, hard rock, and bare land areas have lower

permeability than vegetation areas. Each land use type impacts groundwater retention differently based on evapotranspiration, soil erosion, and runoff (Chotpantarat and Boonkaewwan, 2018; Saravanan et al., 2018). The land use factor is routinely developed through aerial imagery and field surveys (Avtar et al., 2010; Basavaraj and Nijagunappa, 2011; Magesh et al., 2012; Mandal et al., 2016; Shekhar and Pandey, 2015). Soil type directly affects groundwater infiltration (Sun et al., 2018) and is a necessary part of defining groundwater recharge (Oh et al., 2011). Mollinedo et al. (2015) reported that a region's water holding capacity depends on the type, texture, and depth of soil. Díaz-Alcaide and Martínez-Santos (2019) indicated that gravelly and sandy soils are high percolation materials, whereas silty and clayey soils allow the least infiltration. Meanwhile, moderate infiltration is related to loamy soils and fine sand. Soil data are typically collected from available soil maps prepared by the land management department of a nation (Agarwal and Garg, 2016; Kumar et al., 2016; Oh et al., 2011; Pinto et al., 2017). NDVI is also an indicator used in forecasting groundwater storage and aquifer production of a study area (Chen et al., 2018; Naghibi et al., 2017b; Nampak et al., 2014). Areas with a high NDVI index reflect better GWP than those with a low NDVI index (Nampak et al., 2014). NDVI data are usually computed from satellite imagery sources in previous studies.

2.3.1.6. Aquifer-related parameters

In addition to geological-related factors for the formation of consolidated /unconsolidated aquifers, the other characteristics of aquifers, such as aquifer thickness, aquifer resistivity, artesian pressure, groundwater depth, and groundwater quality also influence GWP. In the groundwater study, the thickness of an aquifer also aids in defining potential zones, which is demonstrated through the thickness of weathered materials (Razandi et al., 2015). Shekhar and Pandey (2015) indicated that areas with thick weathering have more groundwater than those with thin weathering because groundwater is prevalent at the base of the weathered zone, wherein rocks

have broken down into sand size and large fragments. In addition to aquifer thickness, aquifer resistivity is also used to determine GWP in some studies (Aizebeokhai et al., 2010; Jha et al., 2010; Muchingami et al., 2012). A high resistivity value indicates low GWP and vice versa (Muchingami et al., 2012). From the viewpoint of the aquifer classification (confined/unconfined), artesian pressure is also a decisive factor in determining potential locations of groundwater. The confined aquifers occur in wide and thick permeable formations with low artesian pressure, which yields low amounts of groundwater, whereas the aquifer systems occur in thin permeable formations with high pressure, thus yielding appreciable amounts of groundwater (Da Lio et al., 2013). Groundwater depth is also one of the key indicators in determining the GWP of a well or borehole location. A large groundwater depth results in a large GWP (Machiwal et al., 2011). In addition, groundwater quality reflects the GWP of an area. Oikonomidis et al. (2015) used nitrate concentration in groundwater and GWP index to delineate GWP and non-GWP zones for domestic uses in Central Greece in the region of Thessaly. Their results indicated that groundwater in the karst area has good storage and low nitrate concentration considering a good GWP zone. By contrast, groundwater in the agricultural activity regions with high nitrate concentration is considered a poor GWP zone despite its high yield. Gopinath et al. (2019) indicated that the intrusion of seawater in eastern coastal India increases dissolved mineral concentrations (such as sodium and chloride) in aquifers due to intensive pumping rates. This phenomenon reduces GWP in areas facing seawater intrusion problems. Furthermore, Chotpantarat et al. (2020) used hydrochemical characteristics, including the stable isotopes ^{18}O and ^{2}H, to assess the origin of the groundwater and applied principal component analysis to determine sources of nitrate contamination in Phetchaburi Province, Thailand. Their result showed that nitrate concentration mainly originated from the agricultural regions, reducing the GWP area in this region. Boonkaewwan et al. (2020) investigated the mechanisms of arsenic (As) release in coastal alluvial aquifers in Rayong Province,

Thailand. They found that arsenite (As^{3+}) is mainly released in the deep coastal aquifer groundwater due to reducing conditions. Thus, groundwater treatment is necessary for some areas before drinking. As previously mentioned, in addition to the groundwater quantity, the assessment of the GWP area should also consider groundwater quality.

2.3.2. Model techniques

Approximately 59% of the 91 publications (Table 2.1) used a model or method to develop and assess GWP. The remaining 31% used two or more models. The current study classified the models into three main groups: statistical, machine learning, and hybrid/ensemble models (Fig. 2.6). The model categorization was complicated, and many revisions were made until the final categorization was realized; this categorization was subjective and based on expert judgment. Fig. 2.6 shows the information regarding the popularity of models used in 91 publications. The database of this study revealed the use of 60 models from 2010 to 2020; the figure is considerably large. Models MCDA-AHP, FR, EBF, WOE are the models with the most occurrences in the statistical group, while RF, LR, BRT, SVM models are used a lot in machine learning group. In this section, the study looks into the characteristics and roles of these popular models.

Fig. 2.6. Popularity of GWP models from 2010 to 2020

2.3.2.1. Common statistical models

a. Multi-criteria decision analysis-Analytical hierarchy process (MCDA-AHP) model

MCDM-AHP is a model that uses associated factors to assess a problem (Saha, 2017). The related factors are assigned a weight corresponding to their importance in observing issues (Kumar et al., 2016). The MCDM-AHP analyzes the weight of related factors to the main variable. The variables in this model are compared considering their effect on each other on a scale of 1–9. The MCDM-AHP model aids hydrogeologists in identifying and decentralizing groundwater influencing factors in the GWP field. The advantages of this method lie in its fast results and minimal errors in choosing well locations. In addition, the MCDM-AHP model permits the modification of the factor weight to satisfy different areas. However, the MCDM-AHP model depends on the opinion of experts in setting the criteria weight; thus, the model result is subjective (Singh et al., 2018). The MCDM-AHP model is often

combined with GIS tools, verifying the geographical regions that agree with GWP conditions. The MCDM-AHP model is typically used to evaluate GWP comprising the following four steps. Step 1: Establishing the factors related to the GWP and creating an important hierarchy for factors. Step 2: Classifying and comparing the factors influencing groundwater. Step 3: Calculating the weighted value of the influencing factors. Step 4: Validating the consistency of the pair comparisons. In the MCDM-AHP, influence factors are combined into an indicator to evaluate outputs (Andualem and Demeke, 2019).

In recent years, hydrological studies have commonly applied the MCDM-AHP model. Elmahdy and Mohamed (2014) delineated GWP zones using the AHP model in the Al Dhaid area, United Arab Emirates. The model was successful in integrating thematic layers and assessing their weights. Selvam et al. (2015) created GWP and recharged zones in the Tuticorin districts of Tamil Nadu, India, using the MCDM technique. The variables used included soil, precipitation, slope, drainage, lithology, land use, and lineaments, where lithology and land use were integrated with the highest weight. Jenifer and Jha (2017) used the MCDA-AHP model for mapping GWP zones in the state of Tamil Nadu of India based on the assigned weights of thematic layers according to the suggestion of experts. The AHP model provided accurate predictions. The groundwater yield was applied to test the validity of the AHP model. Andualem and Demeke (2019) incorporated the MCDM technique and GIS technology to delineate the GWP zones in Upper Blue Nile Basin, Ethiopia. The thematic layers were assigned on the basis of the standardization with effective potential. The results indicated that GWP was categorized into the following five zones: excellent, very good, moderate, poor, and very poor potential fields. Simultaneously, Arabameri et al. (2019) found that the AHP model was more effective than the random forest model by using yield data and thematic layers to assess GWP.

b. Frequency ratio (FR) model

The FR model is applied mainly in landslide research. This model is utilized to assess and predict the landslide locations in a given region (Choi et al., 2012; Jaafari et al., 2014; Pradhan, 2010). In recent years, some hydrogeologists have used the FR model to research groundwater resources (Falah et al., 2017; Guru et al., 2017; Mousavi et al., 2017; Termeh et al., 2019). The FR model is a bivariate statistical algorithm used to compute the occurrence probability rate of a factor (Razandi et al., 2015). Thus, the FR ratio is the rate of non-occurrence or occurrence of a given characteristic. Regarding advantages, the FR is a simple method used to compute the probabilistic connection between the environmental factors and GWP. This method helps discover the spatial correlations between GWP and environmental factors and simplifies the quantification and comprehension of each environmental factor by the model. The FR model is directly employed in a GIS (Oh et al., 2011), and its outcomes are uncomplicated (Falah et al., 2017). However, the FR model depends on the technique-related elements. The adjustment of input parameters will impact the output map (Oh et al., 2011). In reality, the construction of the FR model is conducted on the basis of the observed interactions between each groundwater conditioning factor and the distribution of groundwater well/spring sites.

In the last decade, many studies have conducted GWP maps utilizing the FR model. Ozdemir (2011b) discovered that the FR model was better estimators than Weights of evidence (WOE) and Logistic regression (LR) in mapping GWP in the Sultan Mountains, Turkey. Oh et al. (2011) used the FR model to create the GWP map in the area of Pohang City, Korea. They surveyed 83 well locations; the well points at 55 locations were used for training the model, and the well points at 28 locations were selected for the model test. The training cases were utilized to predict the groundwater appearance availability based on thematic layers, including soil, lineament, geology, and topography data. Their result indicated that soil texture had

the most impact on GWP, whereas ground elevation had the least impact. Razandi et al. (2015) compared the efficiency of the FR, AHP, and gradient boosted decision tree (DT) models in Varamin Plain, Iran, with the areas under the curves (AUCs) of 77.55%, 73.47%, and 65.08%, respectively. Guru et al. (2017) applied the FR to explore the impact of six factors (i.e., DD, lineament density, geology, slope, geomorphology, and land use) on groundwater level. The prediction rate of the FR model was 77.23%. According to Termeh et al. (2019), the FR model was used to create the groundwater spring potential map in Alborz, Iran. They surveyed 339 springs; springs at 238 locations were used for training the model, and springs at 101 locations were selected for the model test. The training cases were utilized to predict the groundwater appearance availability based on thematic layers, including land cover, lithology, DD, fault density, topographic wetness index, plan curvature, slope direction, altitude, soil, precipitation, distance from the river, distance from the fault, profile curvature, slope length, and slope angle.

c. Evidential belief function (EBF) model

The EBF is a bivariate statistic model based on the combination rule to determine spatial integration (Naghibi and Pourghasemi, 2015). The structure of an EBF model includes degree of plausibility (Pls), degree of uncertainty (Unc), degree of disbelief (Dis), and degree of belief (Bel) (Nampak et al., 2014). The Pls and Bel are respectively upper and lower levels of belief with a basic probability assignment function. The Unc is equal to Pls-Bel, which is a lack of understanding evidence of a proposition. The Dis is a value of 1–Pls that ranges from 0 to 1, which is the belief of a false proposition. The relative flexibility of EBF is used in the GWP study to admit uncertainty of spring occurrence from the available evidence (Naghibi and Pourghasemi, 2015). However, the FR requires a considerable amount of information to reduce uncertainty.

Recent GWP projects indicate that the EBF provides a reasonable solution for conceptual methods. Manap et al. (2014) applied the EBF model to investigate groundwater productivity at Langat basin area, Malaysia. The input database included the potential of 12 environmental factors: rainfall, soil, NDVI, land use, lineament density, lithology, DD, TWI, SPI, curvature, slope, and elevation. The main goal was to delineate groundwater productivity zonation and demonstrate the value of the EBF model. Naghibi and Pourghasemi (2015) compared the performance of the EBF to that of the Boosted regression tree (BRT), CART, Random forest (RF), and GLM models using land use, lithology, fault density, river density, distance to faults, distance to rivers, TWI, SPI, slope length, profile curvature, plan curvature, elevation, slope aspect, and slope angle as the environmental factors to map GWP in Chaharmahal-e-Bakhtiari Province, Iran. The assessment output indicated that the performance of the EBF was the lowest among the models. Tahmassebipoor et al. (2016) investigated the GWP with the EBF and WOE approaches. The input factors were utilized as precipitation, lineament density, slope percent, TWI, curvature, elevation, DD, soil texture, distance from river, land use, and lithology. The result showed that the AUC of the EBF and WOE was 83.7% and 78.2%, respectively. Kordestani et al. (2019) simulated GWP in Chaharmahal-e-Bakhtiari Province, Iran by combining the EBF and the BRT, namely the EBF-BRT model. The inputs included topographical and hydrogeological factors. The results revealed that the EBF-BRT was more accurate than the EBF. In another research project, Termeh et al. (2019) compared the ANFIS-GA (genetic algorithm), ANFIS-BBO, ANFIS-SA (simulated annealing), FR, EBF, and entropy models for predicting GWP in Booshehr plain, Iran. They concluded that the ANFIS-GA had superior performance to other models.

d. Weights of evidence (WOE) model

The WOE technique is a method that uses phenomena and events to prove a hypothesis (Al-Abadi, 2015; Falah et al., 2017; Oikonomidis et al., 2015; Ozdemir,

2011b). This method evaluates the weight of variables based on the probability of a phenomenon or event occurring on each variable. This model is derived from the weight of evidence method used in Bayesian statistics. In GWP studies, the WOE estimates the weight for the absence or presence of each subclass of environmental factors based on the absence or presence of representative parameters, such as groundwater yield and spring/well/borehole occurrence, within a given region (Al-Abadi, 2015; Falah et al., 2017; Oikonomidis et al., 2015; Tahmassebipoor et al., 2016). The advantages and disadvantages of the WOE lists are as follows. On the one hand, the WOE helps eliminate the subjectivity of weight and easily sets up the GWP maps. On the other hand, the WOE depends on the number of conditioning variables to set the weight of evidence and require the independence of data input (Tahmassebipoor et al., 2016). Therefore, a requirement for studies using the WOE is to determine the accurate number of input variables, which helps obtain the correct weight of variables in the model.

While researching groundwater productivity potential mapping, Lee et al. (2012) analyzed the correlation of hydrogeological factors with specific capacity (SPC) and transmissivity (T) data. SPC and T were collected at 83 and 81 well locations in Pohang city, South Korea, respectively. SPC and T data were subdivided in a 70:30 ratio to train and validate the output. They concluded that the WOE model's accuracy was 71.20%. The map was generated by the WOE model with relatively accurate groundwater resource exploration. Al-Abadi (2015) applied the WOE model to demarcate GWP in the Missan and Wasit governorates, Iraq. This study utilized data from 143 borehole locations and eight variables affecting GWP, including depth of borehole, aquifer type, distance to faults, distance to roads, LU, geology, slope, and altitude. The results of the GWP map were divided into areas with very high potential (15%), areas with high potential (23%), areas with medium potential (32%), and areas with low potential (30%). Falah et al. (2017) investigated the capabilities of

the WOE, GAM, FR, and SI with different combinations of land use, soil, lithology, distance to faults, fault density, topographic wetness index, distance to rivers, DD, slope aspect, slope angle, plan curvature, altitude, and 6439 springs to predict GWP in Lorestan Province, Iran. Their results showed that the accuracy of the GWP delineation map from the WOE was lower than that of other models. Chen et al. (2018) developed a novel ensemble WOE with LR and functional tree models to map groundwater spring potential in Shaanxi Province of China. They found that the ensemble model is more satisfactory compared with individual models. Falah and Zeinivand (2019) applied the WOE model to compare it with an FR model in the assessment of groundwater distribution in Lorestan province, Iran. They used 212 springs, in which 140 springs were training datasets and 72 springs were test data. Their output data showed that the predictions of the WOE model are less accurate than those of the FR model.

2.3.2.2. Common machine-learning models

a. Random Forest (RF) model

The random forest (RF) is a predictive model that uses many DTs to forecast the appearance of an event, in which each DT is generated from resampling. In 2001, Breiman (2001) was the first author to apply the RF algorithm to classify his data. Then, the RF was further developed and utilized in many science fields. According to Rodriguez-Galiano et al. (2014), regression trees and classification trees could be considered part of the RF model. A regression tree mission involves determining the limitation characteristics and conditions for the formation of a node in a DT, while the function of the classification tree is to classify the data that have not been previously categorized into appropriate groups or classes. The RF model yields very high accuracy results but does not show the operating algorithms because of the complex structure of this model (Rodriguez-Galiano et al., 2014). The output of RF

can be classification or regression. The RF has pros and cons compared with other machine learning models. Regarding advantages, the RF can work with complex data, decrease the overfitting and variance issues, and is remarkably stable. Meanwhile, the disadvantages of the RF lie in its large and powerful computer resource requirement, long training process, and its assignment of the number of variables and trees in the model.

In the area of GWP with the RF, Zabihi et al. (2016) assessed two GWP maps created with RF and MARS for Boujnordm, North Khorasan, Iran. Fourteen explanatory factors and 234 spring locations, including 176 cases, were applied to train the model, and 58 locations were employed to create the GWP. The result showed that the AUC of RF is lower than the AUC of MARS. Similarly, Golkarian et al. (2018) compared the RF, C.50, and MARS in forecasting GWP in Mashhad, Razavi Khorasan Province, Iran. The input variables were lithology, land use, faults density, rivers density, distance to faults, distance to rivers, slope length, TWI, profile curvature, plan curvature, slope angle, slope aspect, and elevation. Their outcome indicated that the RF had an accuracy lower than MARS and higher than the C5.0. Naghibi et al. (2018) compared the KNN, ANN, PDA, QDA, LDA, Support vector machine (SVM), MARS, BRT, and RF models for GWP assessment in Iran. Lithology, land use, fault density, river density, distance from faults, distance from rivers, TWI, SPI, slope length, slope curvature, plan curvature, slope aspect, and altitude were utilized as the input factors of these models. They indicated that the RF was the best among all models in mapping GPW, and four factors with the most impact on GWP included profile curvature, plan curvature, slope, and elevation. Naghibi et al. (2019) simulated GWP by using the RF, CART, BRT, EBFTM, and RTF. They used 13 input factors, including lithology, land use, fault density, distance from faults, river density, distance from rivers, TWI, slope length, profile curvature, plan curvature, altitude, slope aspect, and slope angle, and 273 spring locations. They concluded that the

EBFTM had the highest performance, followed by the RF. Pham et al. (2019) indicated that the RF was a useful model for predicting groundwater spring potential; hence, the maps created by this model could be advantageous for groundwater resource exploration and management.

b. Logistic regression (LR) model

The LR model is commonly used in research on water issue prediction. This model has recently been developed in other groundwater-related areas, such as ground subsidence (Oh et al., 2011) and potential groundwater springs (Nampak et al., 2014; Ozdemir, 2011b; Rahmati et al., 2018). The LR is a regression model that finds the correlation between independent and dependent parameters to create the coefficient (Nguyen et al., 2020b). The result of the LR is only 0 or 1 (binary model). The LR is helpful in GWP studies to predict groundwater absence or presence via spring/non-spring specific yield parameters. Similar to other machine learning models, the LR has the advantages of simplicity and linearity and is reliable in data training. The disadvantage of the LR is easily gotten overfitting as the number of features is higher than that of observation. Additionally, the LR requires the dependent variable that must be the discrete number set and cannot process non-linear problems.

Ozdemir (2011b) created a groundwater spring potential map (GWSP) for the Sultan Mountains, Turkey, integrating an LR model with the GIS environment. The validation of the research surveyed 440 springs and 17 spring-related variables. The authors reported that this was the first study on GWP using an LR approach to identify potential regions of groundwater. The coefficient of 17 spring-related variables was calculated using the binary LR method. The results show that the LR model is a suitable tool to estimate GWSP in a given region. Nampak et al. (2014) compared the LR and the EBF in predicting GWP in Malaysia's Langat River catchment. They considered precipitation, soil type, NDVI, land use, lineament

density, lithology, DD, TWI, SPI, curvature, slope, and elevation as input factors and concluded that the EBF outperformed the LR. Rahmati et al. (2018) used the LR with the DT to map GWP in Khorasan Province, Iran and compared it with the SVM and C4.5 models. The inputs to the models were CI, slope, TPI, TWI, RSP, stream density, distance from stream, lithology, distance to fault, aspect, and elevation. The results indicated that the LR had the highest AUC value, followed by the SVM and C4.5 models. Zandi et al. (2016) applied the LR to map GWP. They used 38 springs to train the model with mean precipitation, land use, relative, fault density, fault distance, petrology, and slope. After the training, the output map was tested through the AUC value. They concluded that the LR was an effective model in mapping GWP. From the results, land use was the environmental factor that impact the highest on GWP. Martínez-Santos and Renard (2020) successfully demonstrated the efficiency of the LR model in the Bauole sub-catchment, Mali, using precipitation, land use, soil, NDVI, TWI, slope, DD, lithology, proximity to surface water, lineament density, and lineaments and 1848 borehole yields.

c. Boosted regression tree (BRT) model

The BRT is a combination model between the boosting method and the DT algorithm. The BRT operates by combining and fitting many DTs based on the boosting method to improve the predictive performance. The output of BRT is an average model from many approximate rules, but boosting is conducted in a stepwise procedure (Nampak et al., 2014). The application of BRT in GWP mapping is compatible with classification and regression of a spring/non-spring location or groundwater yield in a given region (Mousavi et al., 2017; Naghibi and Pourghasemi, 2015; Nampak et al., 2014). Considering its advantages and disadvantages, the BRT can work with numerical and categorical data, reduce overfitting, and optimize the different loss functions. However, the real-time database process of the BRT is long because of the large number of trees, and the output can be outliers.

Naghibi et al. (2016) compared the performance of the BRT to that of the CART, RF, and EBFTM in the spatial modeling of GWP. They used 864 spring locations to train and validate these models according to the rate of 70:30. The input factors included fault density, DD, land use, geology, distance to faults, distance to rivers, profile curvature, plan curvature, slope length, TWI, altitude, slope aspect, and slope degree. Their result showed that the BRT was a good model in predicting GWP with an AUC value of 0.898. Mousavi et al. (2017) used the BRT and the FR to delineate GWP zones in the Mashhad Plain, Iran. Lithology, land use, fault density, river density, distance to rivers, distance from faults, TWI, slope length, profile curvature, plan curvature, elevation, slope aspect, and slope degree were applied as input factors. Both models demonstrated satisfactory performances. Their results showed that TWI had the largest impact on predicting GWP, followed by elevation and distance to rivers. Al-Abadi (2017b) compared the BRT with FR and the ant miner algorithm in assessing GWP in the Euphrates River Basin, Iraq. The slope, profile curvature, plain curvature, TWI, SPI, distance to rivers, distance to faults, altitude, lithology, and aquifer type were used as the input data of these models. The result indicated that the GWP map from the ant miner algorithm was more accurate than that from the BRT and the FR. Kim et al. (2019) investigated the capabilities of the BRT, RF, and LR in predicting specific capacity and transmissivity of groundwater productivity in the Okcheon-gun, South Korea, by using the inputs of plan curvature, convergence index, TRI, depth of groundwater, distance from channel network, distance to faults, lineament density, drainage basin, slope length, TWI, valley depth, relative, slope, soil type, land use, and geology. Their results showed that applying the BRT, RF, and LR with specific capacity and transmissivity had high accuracy.

d. Support vector machine (SVM) model

The SVM is a statistical learning model for regression and classification tasks to determine a hyperplane for training data (Rahmati et al., 2018). The SVM

comprises sigmoid kernels, radial basis function, polynomial, and linear (Rahmati et al., 2018). In the GWP, the SVM considers the environmental factors as the x variable and the validation factors (groundwater yield, specific capacity, and spring/non-spring) as the class label. The optimal separating hyperplane helps classify the training data into subsets or classes (+1, −1) (Lee et al., 2018; Naghibi et al., 2017a; Naghibi et al., 2018; Sameen et al., 2019). The advantage of the SVM lies in its effective operation in high-dimensional spaces and in cases where the number of samples is lower than that of dimensions. However, the SVM underperforms when the dataset is large and the noise is substantial.

Naghibi et al. (2017a) tested four different SVM models along with the random forest genetic algorithm optimized (RFGA) and the RF to create the GWP maps in Ardebil Province, Iran. The environmental factors included lithology, land use, faults density, distance to faults, rivers density, distance from rivers, TWI, SPI, profile curvature, plan curvature, slope aspect, slope angle, and elevation. SVML yielded the best result among different SVMs. However, the performance of all SVM models was lower than the RF and the RFGA. Lee et al. (2018) compared the ANN and SVM for GWP prediction in Boryeong City, Korea. They used 27 hydrogeological factors to predict groundwater yield and SPC. Their results revealed that the ANN was better than the SVM in performance. However, the SVM was proposed as a time-effective GWP tool. Guzman et al. (2019) investigated GW for agricultural activities in the southeastern USA. They considered evapotranspiration, precipitation, and groundwater level data as model input. Their results concluded that the SVR was better than the ANN. Sameen et al. (2019) applied the RF, SVM, ANN, DT, SLRF, and voted ANN-RF to map GWP. The applied data included geology, profile curvature, TWI, STI, SPI, elevation, aspect, slope, curvature, plan curvature, distance to stream, land use, and TRI and 85 wells with different groundwater yields to train and test these models. The comparison between the RF, SVM, ANN, DT, self-learning random

forest (SLRF), and voted ANN-RF for GWP mapping indicated that the SLRF was the most suitable model. Panahi et al. (2020) applied the SVM with grid search and genetic algorithms to forecast GWP in the Ajloun and Jerash Provinces, Jordan. The models were applied to improve the predictive performance in GWP using environmental factors, including NDVI, lithology, land use, rainfall, distance to faults, slope length, soil type, SPI, distance to drainage, TWI, curvature, altitude, aspect, and slope angle. Their results indicated that applying the SVM with the genetic algorithm demonstrated more accuracy than the SVM with the grid search.

2.3.2.3. Hybrid/ensemble models

The evolution of computer science, GIS, and remote sensing has helped models overcome the disadvantages and improve predictive performance. Combined or hybrid models have commonly been used in GWP studies in recent years. The combination of statistical methods and machine learning not only solves data problems but also boosts the predictive productivity of GWP. Additionally, the integration of parametric and nonparametric structures from different sources of knowledge is a strong point of the hybrid/ensemble models. This finding has been demonstrated in many recent GWP studies using hybrid/ensemble models. Khosravi et al. (2018) delineated GWP zones in the Koohdasht–Nourabad Plain in Iran by integrating ANFIS with metaheuristic optimization algorithms. They used lithology, soil type, rainfall, land use, distance from river, distance from fault, TRI, TWI, SPI, plan curvature, elevation, and slope as input data. IWO, DE, FA, PSO, and BA algorithms were applied to optimize the ANFIS. Their result revealed that the ANFIS + DE demonstrated the best model performance. Termeh et al. (2019) proposed a combination between SA, BBO, and GA techniques and ANFIS to optimize the GWP map. The relevant input factors include land use, soil, rainfall, lithology, distance to rivers, fault density, distance to faults, DD, TWI, profile curvature, plan curvature, slope length, slope aspect, slope angle, and altitude. A total of 339 wells were

divided with a scale of 70:30 to train and validate these models. The GA was found to be superior to the SA and BBO method as combined with ANFIS for improving the performance of the GWP map. These models also indicated that rainfall, soil, and land use were important factors for GWP in the Booshehr plain, Iran. Pham et al. (2019) investigated the GWP in the Vadodara district, India, by developing hybrid models based on Decision Stump (DS), RF, MB, and BG. Their database included 34 groundwater wells, slope, soil, land use, lithology, river density, precipitation, TWI, plan curvature, aspect, slope, and altitude. They also integrated DS with other models to classify the different ensemble learners. Their result indicated that the dataset with a scale of 50:50 is superior to other ratios, and the performance of all hybrid models was satisfactory. Singh et al. (2022) created a hybrid model based on catastrophe theory (CT) and AHP to analyze the space of GWP in the West Bengal state, India. Their result removed the major limitation of the catastrophe theory in quantitative factors. The accuracy of this hybrid model was 77%. Nguyen et al. (2020b) developed four ensemble models using LR with cascade generalization, random subspace, bagging, and dagging to map the GWP in Dak Lak Province, Vietnam. They used 12 geo-environmental factors, such as geology, land use, soil, river density, precipitation, flow direction, TWI, STI, curvature, slope, elevation, and aspect, and the groundwater yield of 195 well locations, as input data. Their result revealed that the AUC value of ensemble models was higher than that of the single LR model, in which the LR+ dagging demonstrated the best performance.

2.3.3. Validation

2.3.3.1. Validation-related parameters

The past studies (Machiwal et al., 2011; Naghibi and Pourghasemi, 2015; Rajaveni et al., 2017; Razandi et al., 2015) have applied the groundwater measurement parameters, such as specific yield (Sy)/specific storage (Ss), non-spring/spring occurrence,

and groundwater level, to delineate the GWP zones of a given region. Uliana (2005) stated that "specific storage (Ss) is the volume of water released from a unit volume of a confined aquifer for a unit (1 m or 1 ft) drop in piezometric head, and specific yield (Sy) is the volume of water that can yield by gravity drainage from a unit volume of an unconfined aquifer for a unit drop (1 m or 1 ft) in water level; this quantity is equivalent to the effective porosity." Kresic (2010) indicated that spring is a location where groundwater is discharged to the land surface by gravity and water pressure. The groundwater level is the level of saturated water in soil or rock observed in a well/borehole (Maggirwar et al., 2011). These parameters are crucial for classifying the GWP levels of a given area. The GWP studies from 2010 to 2020 have used different parameters depending on reality conditions and approaches to develop and assess the models. Machiwal et al. (2011) indicated that the high GWP is associated with a large groundwater yield. Specifically, the areas with good GWP have a groundwater yield larger than 5 m^3/h, the areas with medium GWP have groundwater yield ranging from 1 m^3/h to 5 m^3/h, and those with low GWP have a groundwater yield of less than 1 m^3/h. Razandi et al. (2015) conducted a study in the Varamin Plain, Iran and found that wells with groundwater yield of more than 40 m^3/h are known as locations with good groundwater productivity. Ozdemir (2011a) used the non-spring/spring occurrence to determine the GWP levels in Aksehir, Turkey. Locations with spring occurrence were reported to have good groundwater productivity. Non-spring/spring occurrence was also used in the research by Naghibi and Pourghasemi (2015) to delineate GWP zones in the Chaharmahal-e-Bakhtiari Province, Iran. Their result revealed that springs rarely occur in regions with low GWP. Rajaveni et al. (2017) indicated that the groundwater level in a well/borehole also represents GWP at a location. They found that areas with low GWP have large groundwater fluctuations, while areas with high GWP have low groundwater fluctuations. Mukherjee et al. (2012) observed data on groundwater levels to verify the GWP map in Kachchh district, India.

2.3.3.2. Validation technique

Validation is a crucial step in assessing GWP models. Validation testing generally verifies forecast performance. Rajaee et al. (2019) indicated that the predicted results increasingly improved. Algorithm development in models has increased, aided by the evolution of GIS, RS, machine learning technologies. Díaz-Alcaide and Martínez-Santos (2019) also argued that the reason for the increased predictive efficiency is the use of an increased number of variables in GWP mapping. According to statistics from the study dataset (91 publications), four main techniques used by hydrogeologists in groundwater research to check the accuracy of a groundwater map in their area of interest are as follows: receiver operating characteristic (ROC), statistical evaluation measures, Kappa index, and root mean square error (Table 2.2). Among these techniques, the ROC is mostly used in GWP studies with model performance evaluation. The choice of the groundwater map validation technique depends on the preference of the operator. Performance evaluation results for common models in this study are shown in Fig. 2.7.
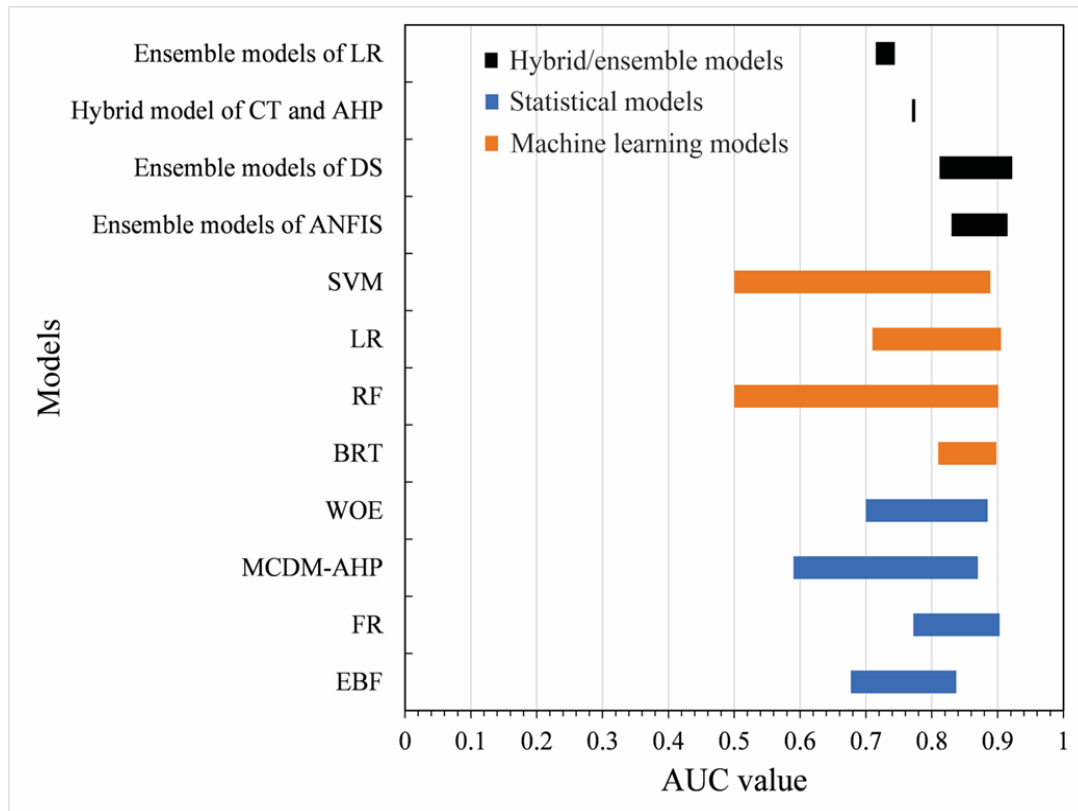
Fig. 2.7. Fluctuation of AUC value in common models in GWP studies

Table 2.2. Validation methods in GWP studies

No	Validation method	Descriptions	References
1	Receiver operating characteristic	The ROC is a curve line graph that describes the relationship between sensitivity (Y-axis) and 1-specificity (X-axis) for every possible cutoff value. The value of ROC is the under-curve area ranging from 0.5 to 1. (0.5-0.6: poor; 0.6-0.7: average; 0.7-0.8: good; 0.8-0.9: very good; 0.9-1: excellent).	(Naghibi et al., 2017a)
2	Statistical evaluation measures	$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$ $$Sensitivity\ analysis = \frac{TP}{TP+FN}$$ $$Specificity\ analysis = \frac{TN}{TN+FP}$$ where TP – true positive; TN – true negative; FP	(Nguyen et al., 2020b)

3	Kappa coefficient	Kappa coefficient is an index to calculate the model reliability in GWP classification. $$Kappa = \frac{Accuracy - P_e}{1 - P_e}$$ P_e is an agreement level that might be generated by chance. The closer the kappa value gets to 1, the better model's performance is.	(Golkarian et al., 2018)
4	Root mean square error	Root mean square is a value that measures the deviations of all samples between the predicted value (X_p) and the actual value (X_a). $$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_p - X_a)^2}$$ The closer the RMSE value is to 0, the better model's performance is.	(Rahmati and Melesse, 2016)

2.4. Current status and challenges for groundwater potential mapping research in the future

This review provides an overview of factors pertaining to GWP mapping from model thematic layers, algorithms, and their accuracy. A GWP map must present the potential areas of groundwater (Elmahdy and Mohamed, 2014). Hydrogeologists usually delineate their potential maps into five levels: very high, high, moderate, poor, and very poor potential regions (Kumar et al., 2014; Mukherjee et al., 2012; Pham et al., 2019). According to the statistics of this review, there are eight topical map layers commonly used by hydrogeologists: geology, slope, land use, soil type, drainage density, lineament density, altitude, rainfall.

Another consideration of GWP map is the data source for mapping thematic layers. For data sources to create thematic layers, two main sources have been commonly employed: satellite data and hydrogeological data. Today, the development of RS technology has provided a large amount of data on the Earth's

surface and weather information, facilitating map creation within a shorter time than that required for field data collection (Díaz-Alcaide and Martínez-Santos, 2019). Thus, although RS data are effective, scientists must have a wide and in-depth knowledge of RS (Oikonomidis et al., 2015). Some satellites with high resolution provide high-quality images; however, this makes the images cost-intensive making their use unfeasible. Another disadvantage of RS data is that satellite images only provide coverage information; information of objects present under cloud cover and treetops is not considered. This requires hydrogeologists to combine information from the field, such as the borehole, well, and hydrogeological data, with RS; field parameters can provide what RS data cannot. For example, well productivity data cannot be collected from satellite images. Additionally, field data always provide a high degree of confidence and validity. However, the collection of borehole, well, and hydrogeological data is time- and labor-consuming. According to current statistics, there are several approaches to establish a GWP map. These methods can be divided into three groups: statistical, machine learning, hybrid/ensemble techniques. The common statistical techniques include MCDA-AHP, FR, EBF, and WOE models. The features of the statistical models are easy to implement a GWP map, their observed variables do not require more than that of machine learning models. In some cases, several authors have only used the weights of the affected variables of groundwater, ignoring the map's accuracy (Gaur et al., 2011; Hashim et al., 2013; Murmu et al., 2019; Singh et al., 2013). Recently, machine learning models have been used extensively by hydrogeologists. The common machine learning models include RF, LR, BRT, and SVM. Big data are required to train and test the machine learning models (Martínez-Santos and Renard, 2020). From basic machine learning and statistical models, hydrogeologists have developed advanced models aimed at creating more accurate GWP maps such as the adaptive neuro-fuzzy inference system (ANFIS) + differential evolution (DE) model (Khosravi et al., 2018), random

forest (RF) + decision Strump (DS) (Pham et al., 2019), adaptive neuro-fuzzy inference system (ANFIS) + genetic algorithm (GA) (Termeh et al., 2019), and logistic regression (LR) + dagging (DLR) (Nguyen et al., 2020b).

An integral part of any GWP study is the validation step (Das, 2019). Statistically, receiver operating characteristic (ROC), statistical evaluation measures, Kappa index, and root mean square error are four main techniques used to evaluate a GWP map. Statistical evaluation measures, Kappa index, and root mean square error are often used to check the accuracy of statistical models, while the AUC technique is often used to check the accuracy of machine learning models and several statistical models. Generally, the effectiveness of the common statistical models ranges from approximately 59.0% to 90.3% while the accuracy of GWP maps produced by common machine-learning models goes from 50.0% to 90.1%. Strikingly, a higher accuracy interval (71.0% - 92.0%) was shown in the hybrid/ensemble models (Fig. 2.7). In addition, some authors have combined the AUC technique with specificity, sensitivity, negative predictive value, and the positive predictive value (Nguyen et al., 2020b; Ozdemir, 2011a; Pham et al., 2019).

The complex changes in climatic conditions and an increase in the demand for clean water pose many challenges for hydrogeologists with regard to groundwater use and planning. In reality, hydrogeologists have conducted extensive researches and pursued various methods and approaches to create the most suitable GWP map for their regions of interest in the past decade. However, GWP mapping research is still characterized by a number of limitations and challenges that need to be addressed in the future.

Thus far, researchers have not found the optimal models required to map the GWP of different areas around the world. There are two reasons for this: the difference between regions in natural conditions and social characteristics, and the discrepancy in the input data. Therefore, it is difficult to conclude which model is the most accurate. Simultaneously, the constant changes in climatic conditions, runoff factors, land-use

coefficients, and population growth directly affect the GWP of a certain region, posing great difficulties for hydrogeologists in mapping GWP.

To address these challenges, future research should focus on the following aspects. First, studies must focus on improving the quality and quantity of the input database to ensure the reliability of GWP maps. At the same time, stronger emphasis must be placed on using satellite images with high resolution to improve the accuracy of classification, which will lead to the improved accuracy of the forecast results. Third, maps should be created based on a combination of machine learning and statistical models. Fourth, researchers need to apply different approaches for each research region to acquire the most suitable research model. Finally, three-dimensional models should be used to obtain an overview of the groundwater system.

2.5. Summary

This chapter aims to review parameters, model techniques, validation methods in groundwater potential field. According to statistics, there are three major model groups used to establish groundwater potential maps. The first model group is a statistic group, including multi-criteria decision making/analytic hierarchy process, frequency ratio, evidence belief function, and weights of evidence. The second model group includes machine learning models, such as random forest, logistic regression, boosted regression tree, and support vector machine. The final group is the hybrid/ensemble models. In groundwater potential mapping studies, 41 thematic layers affect the potential of groundwater. However, hydrological researchers have frequently used eight factors in groundwater potential studies: geology, slope, land use, soil type, drainage density, lineament density, altitude, rainfall. Most previous studies on groundwater potential have used a combination of geographic information system, remote sensing, and machine learning techniques to design the groundwater potential in regions of interest. Data sources are commonly applied to groundwater potential mapping, including satellite, borehole, and geophysical data. The

accuracy of groundwater potential maps produced by common machine learning models ranges from 50.0% to 90.1%, while that produced by common statistical models ranges between 59.0% and 90.3%. Interestingly, hybrid/ensemble models' accuracy interval was from 71.0% to 92.0%. Therefore, the review suggests that statistical algorithms and machine learning techniques should be combined, and thematic layers should be increasingly used in mapping groundwater potential maps to achieve high efficiency.

Chapter 3. STUDY AREA AND RESEARCH METHODOLOGIES

3.1. Study area

3.1.1. Location

The present study focuses on the distribution of potential groundwater in Kanchanaburi province, Thailand. The province is placed in the western central of Thailand, from 99°10'54'' to 99°53'31''E and 13°43'34''to 15°39'46''N (Fig. 3.1). There are approximately 19,483 km^2, of which 61% is a mountain with an elevation from 2 m to 2,028 meters (6653 ft) above sea level. The province has 847.47 km long of the border with neighboring provinces and Myanmar. The region has a tropical savanna climate with significant seasonal variations in temperature and precipitation. Administratively, Kanchanaburi is separated into 13 districts: Bo Phloi, Dan Makham Tia, Huai Krachao, Lao Khwan, Mueang Kanchanaburi, Nong Prue, Phanom Thuan, Sai Yok, Sangkhla Buri, Si Sawat, Tha Maka, Tha Muang, and Thong Pha Phum.
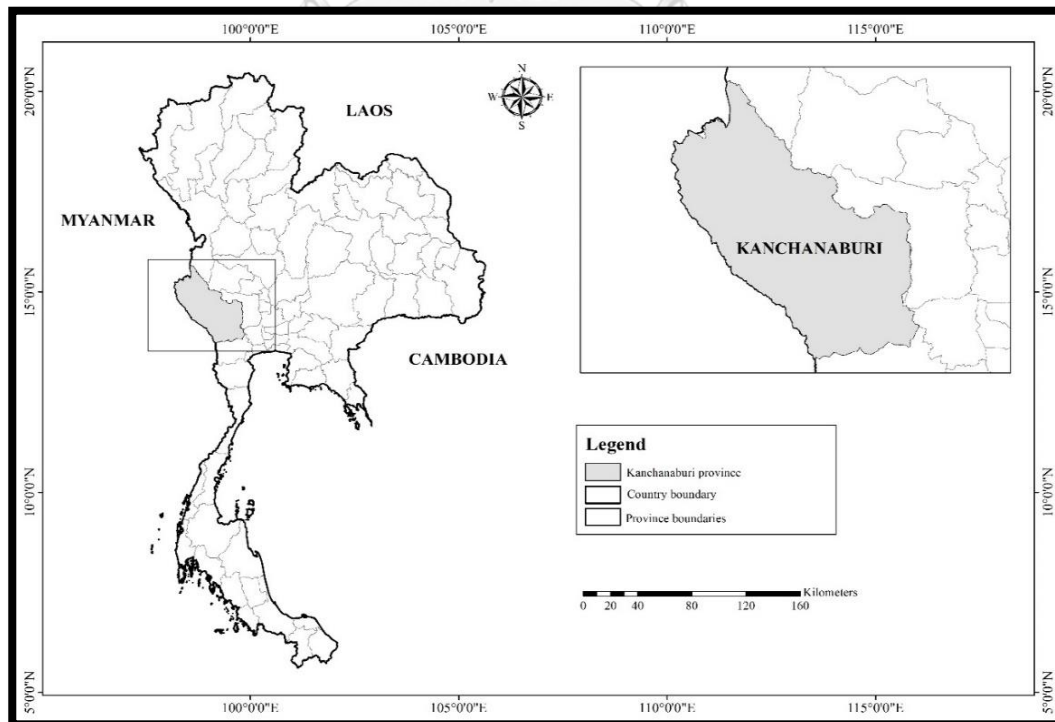
Fig. 3.1. Location of Kanchanaburi province

As of December 31ˢᵗ, 2018, the number of inhabitants of Kanchanaburi was 848,198 and, the population density was around 43.5 people/km^2 (the Statistical Yearbook of Thailand 2018). Four population groups are living in Kanchanaburi, including Thai, Karen and Mon, and Burmese.

Gender (C 2018) (Person)	
Males	420,099
Females	423,081

Age Groups (C 2018) (Person)	
0-18 years	193,643
19-49 years	388,729
50-60 years	145,805
60+ years	120,020
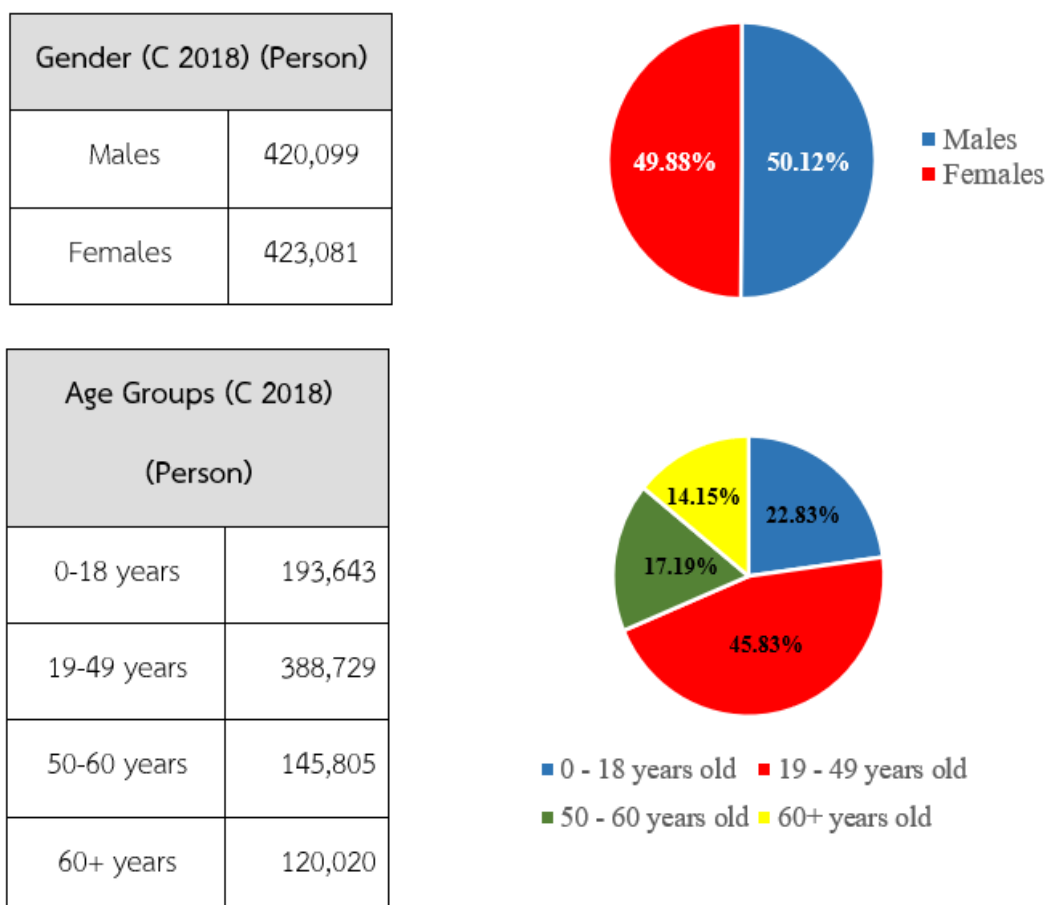
Fig. 3.2. Information about the population structure in Kanchanaburi province

3.1.2. Topography

The terrain of Kanchanaburi province has a very diverse and hierarchical structure. The topography of Kanchanaburi province gradually rises from the East to the West. There are three kinds of terrain in Kanchanaburi: mountains, valleys, and river basins. The mountainous topography is allocated in the North and West of province (in Sangkhla Buri, Thong Pha Phum, Si Sawat, some parts of Sai Yok, Bo Phloi, and Mueang Kanchanaburi districts) with an elevation ranging from 300 to 1800

meters, occupy three-fifths of total land area. The mountainous regions are mainly enveloped by natural and production forests with plant biodiversity.

The valleys area is located in the northeast of the province (in Lao Khwan, Phanom Thuan District, Bo Phloi, and Mueang Kanchanaburi districts) and comprises 20% of the total area. The characteristic of topographic structures is a hillside plain with low hills and groves, and slope direction is from east to southeast, mostly between 30 and 100 meters above sea level. The remaining area is the river basin, mainly located in the south and east of Kanchanaburi province (in Tha Muang. Tha Maka, and some parts of Phanom Thuan districts).
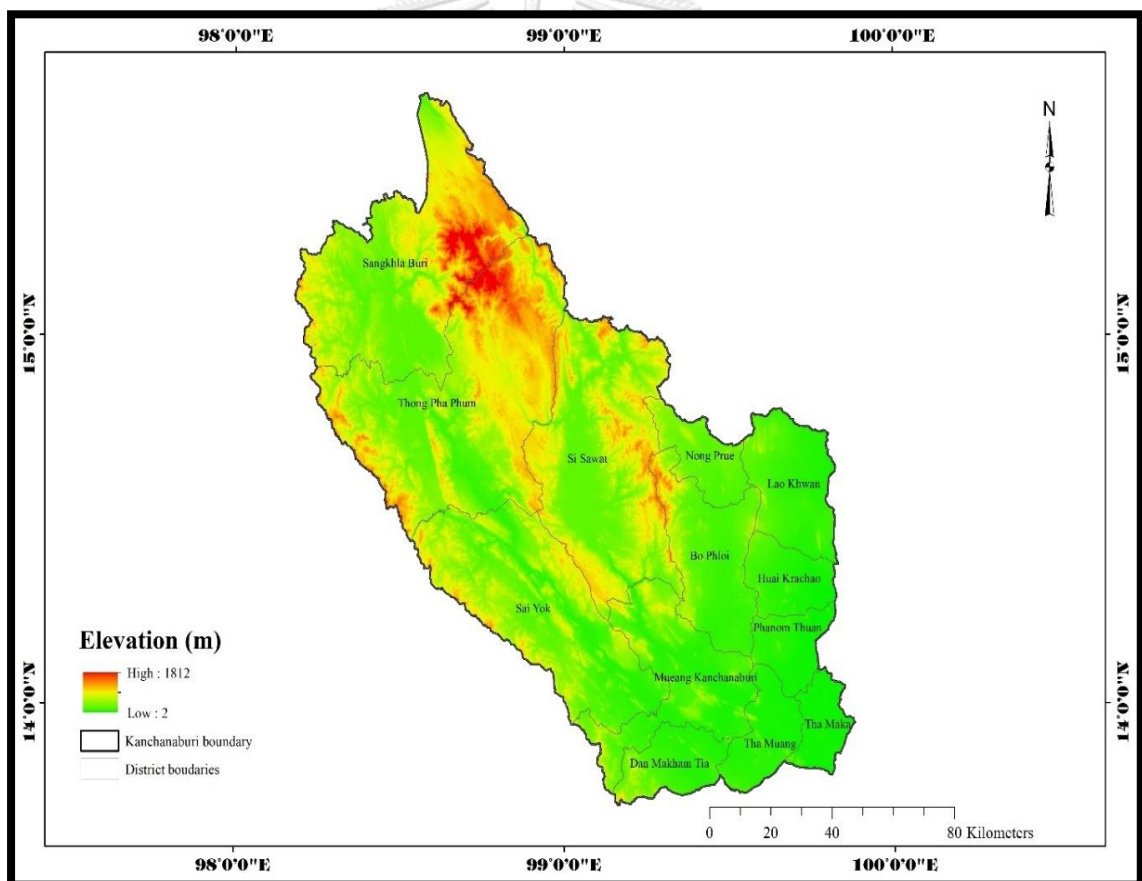
Fig. 3.3. Terrain of Kanchanaburi province in Thailand

3.1.3. Meteorological condition

The atmospheric circulation of Southeast Asian monsoons impacts the climate in

Kanchanaburi. It is the interconversion of climatic characteristics of the mountain and plain Thailand. The northeast monsoon often occurs between November and March; the southwest monsoon usually begins between May and September. The transition period is April and October. The province has two main seasons: summer and winter. In the dry season, the heat lasts a long time; conversely, cold spells occasionally appear in the wet duration. On the report of the Kanchanaburi people's committee about climate, weather features of the province are characterized as follows:

Temperature of the province is very high. It is approximately 38.2°C in the summer and around 31.3°C in the winter. The amplitude of temperature oscillates quite widely. April has the highest monthly mean temperature at approximately 38.2°C, and December is the lowest monthly mean one at just under 19.6°C (https://www.weather-atlas.com).
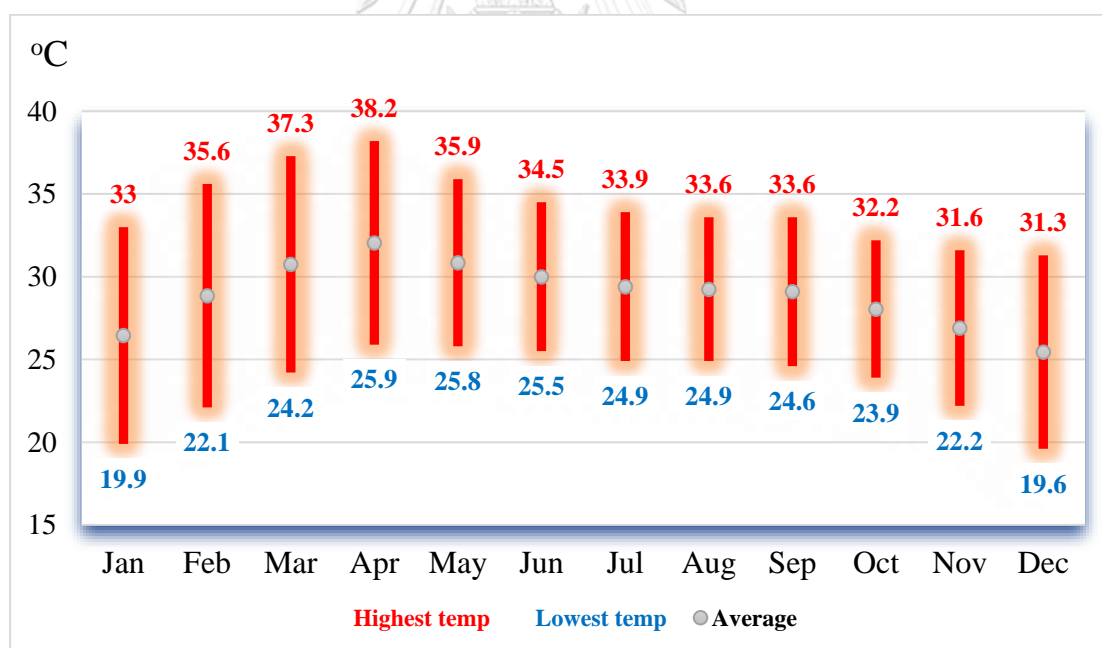
Fig. 3.4. Monthly temperature in Kanchanaburi province (2020)

The annual average relative humidity fluctuates from 61% to 80%. October typically occurs the highest humidity at 80%, and March usually has the lowest humidity at about 61%.
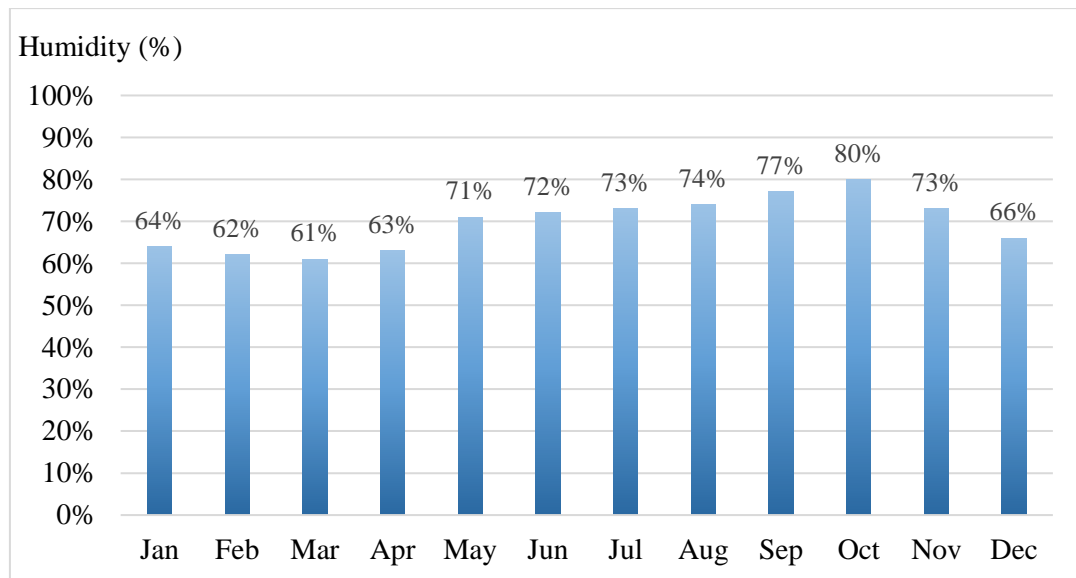
Fig. 3.5. Monthly humidity in Kanchanaburi province (2020)

The annual average precipitation is 1056.4 mm, but it does not distribute evenly in the province. The heavy rainfall is concentrated in mountainous areas. Between May and October, precipitation occupies around 83% of the annual rainfall total. The rest of the rainfall happens from November to April. September is the highest average rainfall of 220.5 mm, while January is the lowest with 3.3 mm.
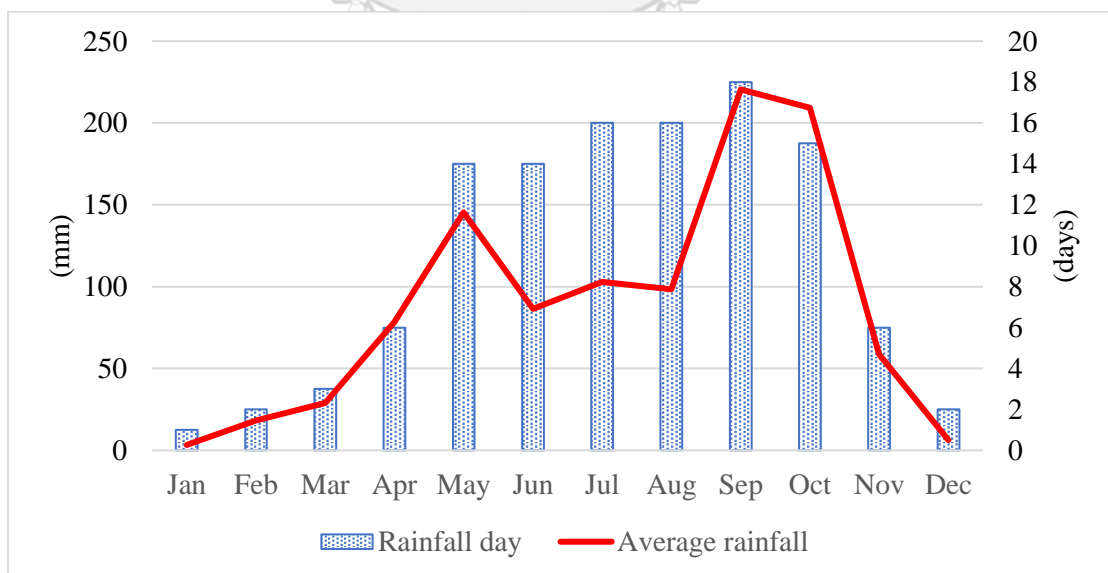
Fig. 3.6. Monthly average precipitation and rainfall days in Kanchanaburi province (2020)

The annual average sunshine and daylight hours are approximately 31,700 hours

and 53,000 hours, respectively. The highest hours of sunshine happen from January to April and December; the average hours of sunshine in these months are around 9 hours per day. Moreover, the lowest hours of sunshine appear from July to September; the average hours of sunshine in these months are approximately 5 hours per day. June and July have the longest days, with 12.9 hours per day; meanwhile, December has the shortest day, with 11.3 hours per day.
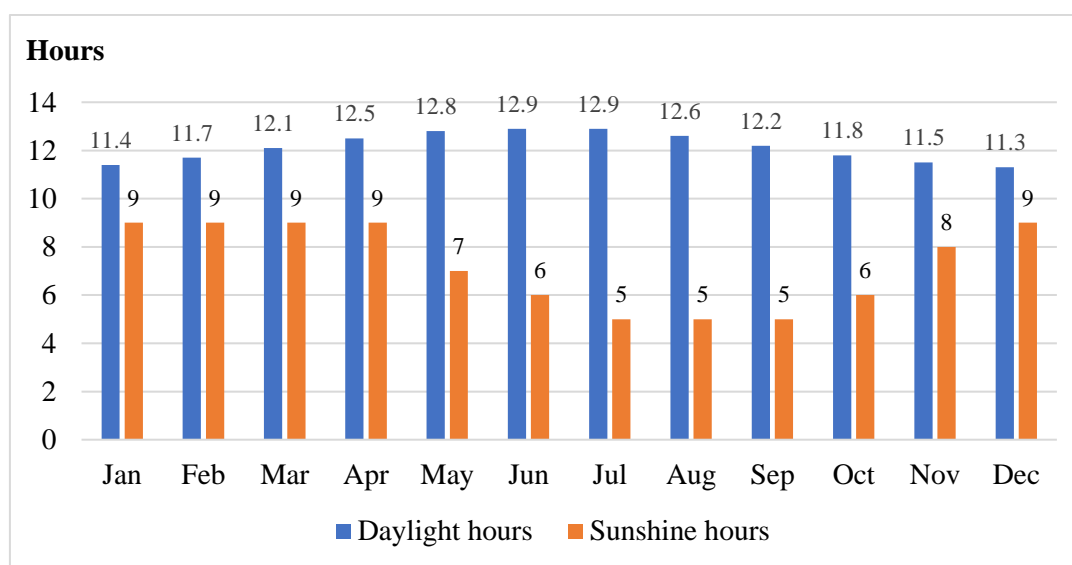
Fig. 3.7. Monthly average sunshine and daylight hours in Kanchanaburi province (2020)

The Southeast monsoon's wind regime begins to blow into the Gulf of Thailand in late May and early June. The southeast monsoon blows through, causing more rain in the period from August to September. In winter, the northeast monsoon blows from the high-pressure area from China. The northeast monsoon is cold and dry. It blows over the whole of Kanchanaburi from October to February next year. The Northeast Monsoon begins to weaken in February, when the East Wind from the South China Sea and the Southeast from the Gulf of Thailand will take its place. In addition to the monsoons mentioned above, the area has occasional storms.

Climate change affects many weather elements such as the number of precipitations, temperature, sunshine duration, etc. Therefore, it impacts the groundwater

quality and quantity in Kanchanaburi. However, the effect of climate on groundwater takes a bit long time (for example, 50 years). According to Srisuk et al. (2016), climate change has degraded groundwater quality in coastal and inland areas in Thailand, such as saline water intrusion. While the analysis by Pholkern et al. (2018) indicated that groundwater resources in Central Huai Luang Basis are at risk due to climate change.

3.1.4. Hydrologic condition
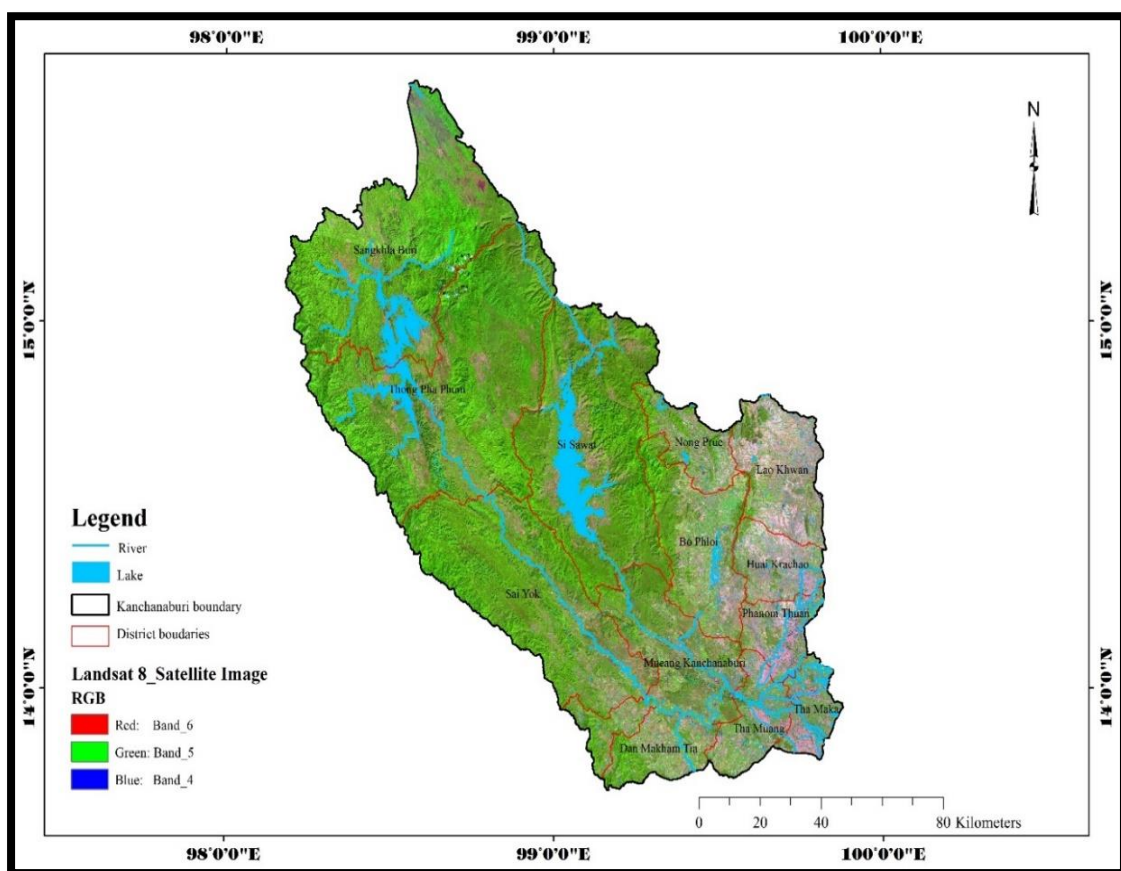
Fig. 3.8. River systems and waterbody in Kanchanaburi province

Four river systems pass the Kanchanaburi province's territory, namely Khwae Noi, Khwae Yai, Mae Klong, and Lam Taphoen rivers. These rivers originate from the Khwae Noi river systems in Thong Pha Phum district and Thong Pha Phum district. Khwae Noi converges with Khwae Yai to form Mae Klong river in Pak Phraek ward, Mueang Kanchanaburi districts. Lam Taphoen river flows through the Eastern Plain of Kanchanaburi province and meets the Khwae Yai River at Tha Sao in Sai Yok district.

3.1.5. Soil resources

Kanchanaburi province has suitable soil conditions for cultivating crops such as sugarcane, maize, cassava, and pineapple. According to the survey of the land development department of Thailand (2007), the soil condition in Kanchanaburi is mainly formed from sedimentary rock by decomposition of igneous rock, granite, granodiorite, shale, and quartz phyllite. There are 33 soil group types found in Kanchanaburi province. Soil group 62 occupies 10,334.81 km^2, followed by soil group 33 (1029.55 km^2), and soil group 29 (994.26 km^2). Soil groups 31 and 48 have a total area of 935.54 km^2 and 841.98 km^2, respectively.

+ Soil group 62: This type of soil covers mainly hilly areas. It appears at places with a slope of more than 35%. The characteristic is deep and shallow. Soil texture and fertility vary according to the original rock, such as crushed stone, lump rock, or slate. This type of soil should not be used for agricultural purposes because of ecosystem conservation.

+ Soil group 33: Soil texture has brown or reddish-brown. It lies deep in the soil, contains mica or lime, is formed from sediment materials, and is found along rivers, hills, and mountains. This soil type usually appears in regions with a slope of about 2-12%. It has drainage from good to moderate; therefore, the groundwater level is more profound than 1m all year round. The soil group has moderate natural fertility. The topsoil layer has a pH in the range of 6.5-7.5.

+ Soil group 29: Soil texture is clay and delicate, with yellowish-brown or reddish colors. It originates from river sediments or degraded soils. This group is found in hills. It appears in areas with a slope of about 3-25%. It is deep and well-drained soil. Natural fertility is relatively low; the pH of soil ranges around 4.5-5.5. This type of land is currently used to grow different types of rice, crops, and fruit trees.

+ Soil group 31: Soil texture is clay. The feature is fertile. It has brown, yellow, and red colors due to the decomposition of many types of rocks. It is found in undulating and deformed regions. This group appears in regions with a slope of about 3-20%. This type of soil is deep and medium well-drained soil. The groundwater level is usually more than 1m deep. The natural average pH is 5.5-6.5. Currently, this type of soil is used to grow rice and various fruit trees.

+ Soil group 7: Soil texture is clay, has brown or gray-brown colors, and is formed by river alluvium. It is deep soil and poorly drained. It is found in flat and relatively smooth areas. The water level is 30-50 cm deep in the rainy season for 3-4 months. This group has naturally fertile and moderate fertility features. The pH ranges from 6.0-7.0.

3.1.6. Geological condition

Kanchanaburi is Thailand's westernmost province. It has a lot of high mountain ranges. The research study's geology consists mainly of metamorphic, igneous, and sedimentary rocks ranging in age from Precambrian to Quaternary. Fig 3.9 depicts the geology in Kanchanaburi province.
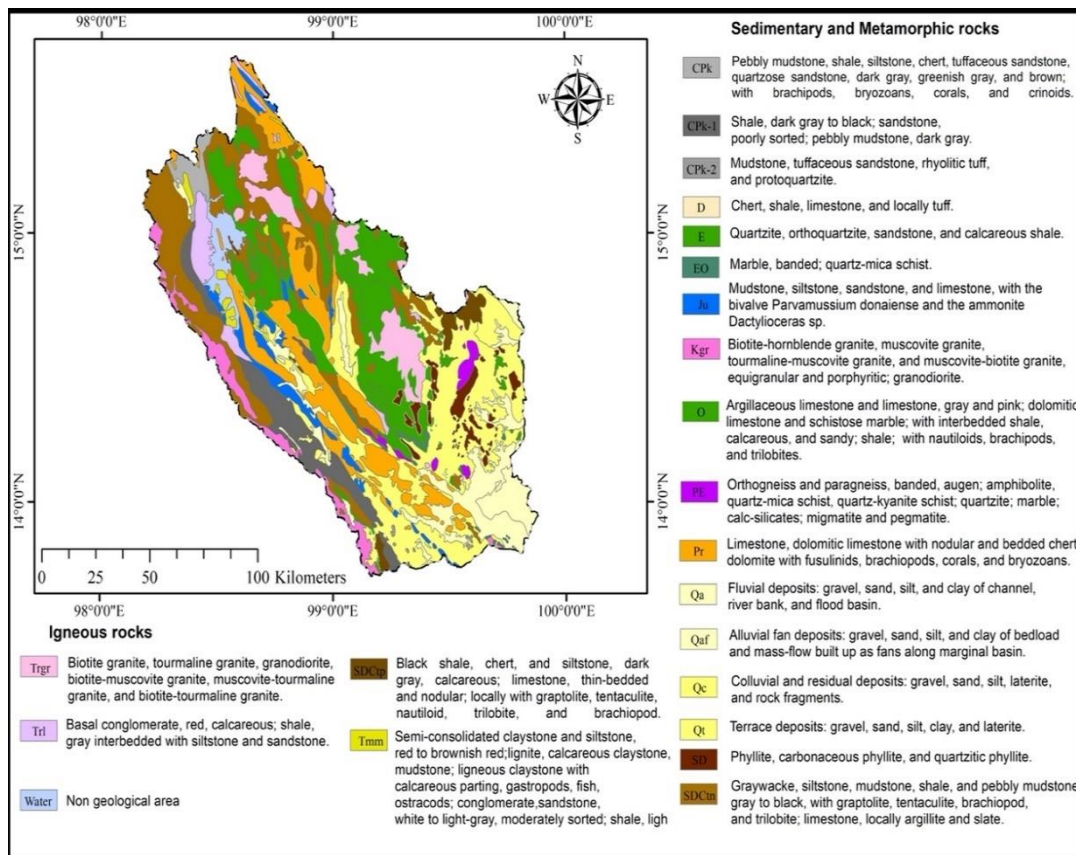
Fig. 3.9. The geological map of Kanchanaburi province

3.1.6.1. Geological structure

The research area's structures are complex with folds and faults. The folds may be seen in various spots in Kanchanaburi from both field and satellite imagery (Songmuang et al., 2007). Most fold axes are in the northwest-southeast direction (Bunopas, 1976). Fold structures may be seen in a variety of rocks. It could be anticlinal and synclinal folds. The anticlinal and synclinal folds have been observed in the Ordovician and Permian limestones in the Khao Leam region. According to (Songmuang et al., 2007), tight and recumbent folds may be found in Ordovician to Devonian strata, mainly in the western section of Kanchanaburi city and the Three Pagoda Fault.

Many faults appear in western Thailand grouped into the Sri Sawat Fault and the Three Pagoda Fault (Chuaviroj, 1991; Pailoplee et al., 2009). The Sri Sawat Fault, which runs north-south, has a curved pattern with multiple smaller fault sets at its

southern end. The Sri Sawat Faults begin in Kanchanaburi Province's Sri Sawat district and the northern section of Bo Ploi district. In a northwesterly direction, this fault also flows through the Khwae Yai River. The Three Pagoda Fault deviates from the Mae-Ping Fault to the north, cutting Paleozoic to Cenozoic lithological units. Fig. 3.10 displays the active faults in Kanchanaburi province.
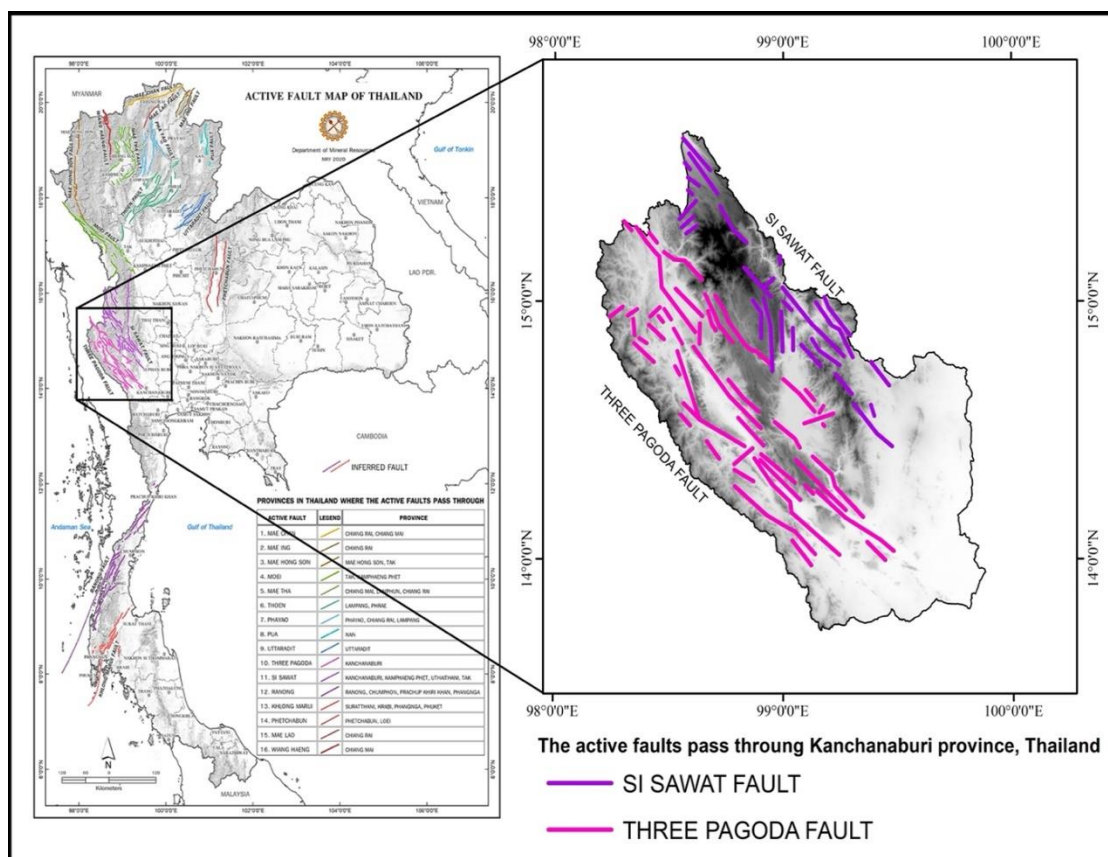
Fig. 3.10. The faults pass through Kanchanaburi province

3..1.6.2. Geological Setting

Kanchanaburi's geology comprises several faults that run in different directions, primarily northwest-southeast. These faults cut through various lithologies. The youngest unit consists of narrow and long strips of Quaternary deposits filled with laterites, silts, sands, and gravels. The Triassic rocks of the Chong Khab formation lie unconformably under the Quaternary layers. The Daonella sp, Halobia sp and sandstone interbedded with limestone make up the Chong Khab formation. The Permian Tha Madua sandstone,

which includes heavily bedded red sandstone and white quartz sandstone, underpins the Chong Khab formation. The Sai Yok Limestone is almost equivalent in age to the Tha Madua sandstone. This Permian-aged deposit consists of massive and bedded limestones with fusulinids, brachiopods, pelecypods, and bryozoans. The PermoCarboniferous-aged Khaeng Krachan formation underlies the Permian-aged Sai Yok formation. The Bo Ploi and Kanchanaburi formations, which are Silurian to Devonian in age, comprise pebbly mudstone, gray sandstone, and dark-gray shale. The Thung Song includes banded argillaceous limestone, argillite, and quartzite with Ordovician cephalopods is older than the Silurian-Devonian rocks. The Tarutao (or Chao Nen Group) of the Cambrian age is the earliest lithologic unit in the study area (Songmuang et al., 2007).

3.1.7. Hydrogeological characteristics

Groundwater sources in Kanchanaburi provinces can be divided into two main groups: Unconsolidated aquifers and Consolidated aquifer.

- Unconsolidated aquifers: Groundwater in unconsolidated rocks differ in both quality and quantity, depending on the type of sediment, biological characteristics, and structural characteristics of sediment accumulation. The areas in the South and East of Kanchanaburi are low-lying areas with a source of groundwater that can be utilized. Groundwater can be withdrawn to use at a rate of 30-50 m^3/hour.

- Consolidated aquifer: Groundwater is stored in fractured materials such as granite, basalt, and limestone. The Western part of the Kanchanaburi province is mainly metamorphic rocks. Therefore, groundwater is shallow, less than 2 m^3/hour (Sangkhla Buri, Sai Yok, Si Sawat, and Bo Ploi districts). Meanwhile, in places with low terrain (Mueang Kanchanaburi, Sai Yok, and Thong Pha Phum districts), the groundwater supplies above the threshold of more than 50 m^3/hour.

In addition to limited groundwater resources, groundwater quality is another

issue in the study area. The total mass of matter is considered the overall chemical characteristics of groundwater. Total dissolved solid (TDS) is a term that uses to describe the total concentration of dissolved materials in water. The optimal value of total dissolved solids (TDS) is 600 mg/l and the maximum allowed is 1,200 mg/l.

TDS in groundwater in Si Sawat, Thong Pha Phum, and Sangkhla Buri districts was less than 500 mg/L. Meanwhile, TDS in groundwater in Sai Yok, Thong Pha Phum, Lao Khwan, and Bo Phloi districts was around 500-1,500 mg/l.

3.2. Research data

3.2.1. Data about groundwater yield assessment

3.2.1.1. Groundwater yield

Our case study considered eight site-specific variables, including altitude, distance to faults, distance to waterbodies, slope, geology, land use, rainfall, and soil type (Fig. 3.11). In previous studies, these variables were considered to influence groundwater yield and distribution (Abd Manap et al., 2014; Arulbalaji et al., 2019; Mumtaz et al., 2019; Naghibi et al., 2017a; Razandi et al., 2015). This study mostly obtained well-processed data from government agencies or, in some cases , extracted data from satellites using well-defined procedures to make our analysis workflow consistent for future studies (Díaz-Alcaide and Martínez-Santos, 2019). In particular, the altitude and slope layers were calculated based on the Aster Global Digital Elevation Model V003 from the National Aeronautics and Space Administration (NASA; https://search.earthdata.nasa.gov) using ArcGIS pro 2.8 (Khal et al., 2020). Land use and soil type maps were obtained from the Thailand Department of Agriculture. Global satellite precipitation data (CHRS, http://chrs.web.uci.e.du) were collected from the rainfall database to estimate the annual precipitation in Kanchanaburi Province during the time period between 2010 and 2020. The distance to waterbodies was digitized and calculated from Thailand's hydrological system. Geological and fault data were

collected from the Department of Mineral Resources. All spatial input layers were resampled to a 30 m resolution for the overlay analysis. Because the AHP and FR methods require categorical input variables for the analysis, the continuous variables in our study (that is, altitude, distance to faults, distance to waterbodies, slope, and rainfall) were classified into subclasses using the equal interval method.

3.2.1.2. Ground truth data

Our case study used groundwater yield as the ground truth data for GWP model training and map validation. Groundwater yield was collected from 1,601 wells across Kanchanaburi Province in a previous project funded by the Thailand Department of Groundwater Resources (DGR). The groundwater yield for 1,601 wells was partitioned based on the threshold of 10 m^3/h into two datasets: those of wells with a groundwater yield of >10 m^3/h and those of wells with a groundwater yield ≤10 m^3/h (Fig. 3.11). Both datasets were used to train and validate the RF model, whereas only the dataset with a groundwater yield of >10 m^3/h was used in the AHP and FR model training. The datasets were split into two subsets using a ratio of 70:30 for model training and validation, respectively. The detailed experimental designs for individual model buildings are presented in Section 3.3.1.
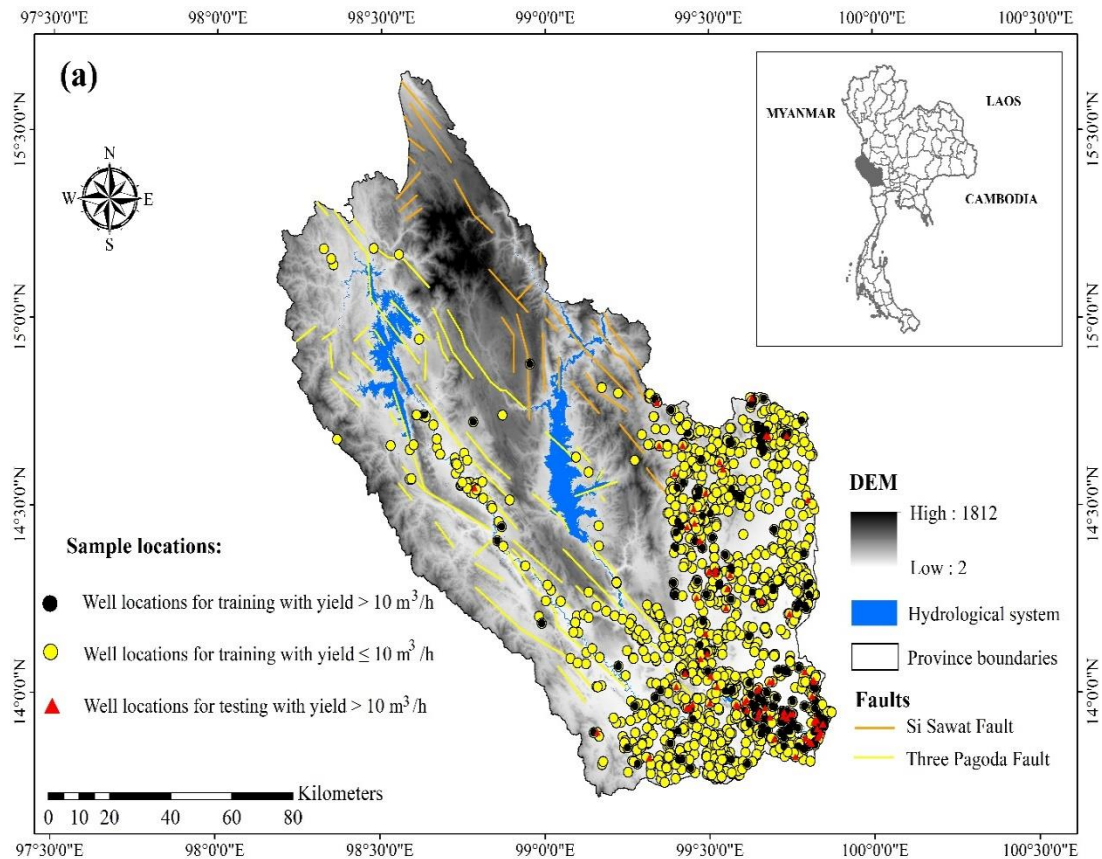
Fig. 3.11. Well positions in Kanchanaburi Province, Thailand

3.2.2. Data about heavy mental contamination risk assessment in groundwater

3.2.2.1. Ni contamination in groundwater

Ni contamination in groundwater is a phenomenon in which the Ni concentration exceeds the permissible level. According to Thailand's groundwater quality standards, Ni concentration in groundwater must not be more than 20 µg/L. For our case study, Ni data of 180 groundwater samples were collected from April to July 2021 and analyzed at Thailand Department of Groundwater Resources (DGR). The groundwater sample locations were displayed using the UTM coordinate system (Fig. 3.12). The concentration of Ni was measured in the laboratory by Atomic Absorption Spectrophotometry-Direct Aspiration method with a detection limit of 10 µg/L. Ni concentration in groundwater samples fluctuates from 10 to 72 µg/L (Fig. 3.13). In 180 groundwater samples, 16 groundwater samples were contaminated with Ni.
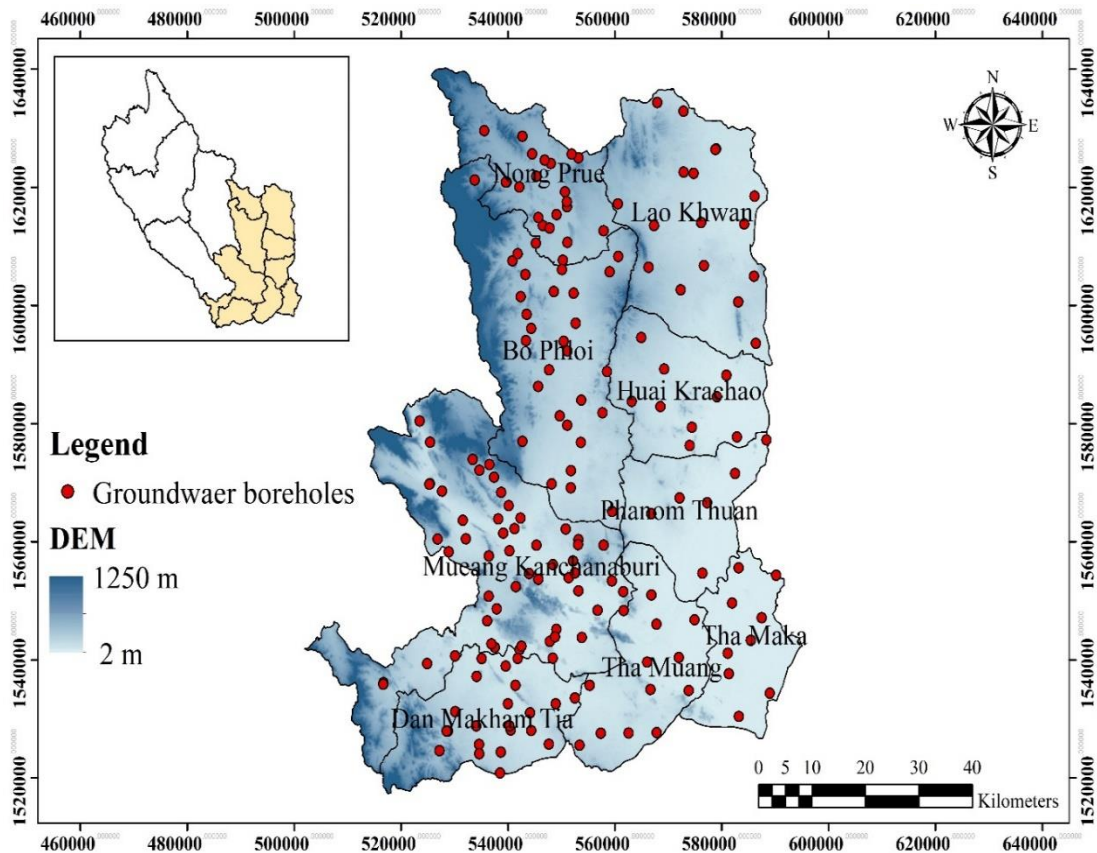
Fig. 3.12. Positions of groundwater samples in study area
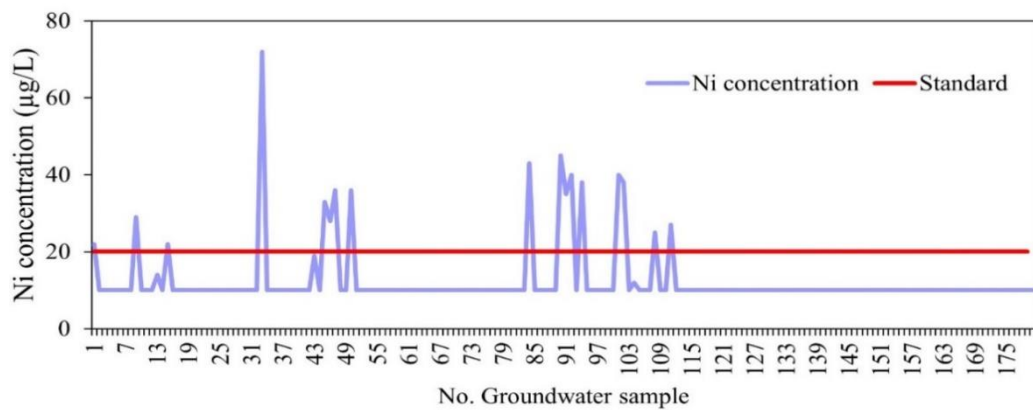
Fig. 3.13. Classification of Ni contaminated and uncontaminated groundwater samples

3.2.2.2. Influencing factors

The groundwater could be contaminated with Ni from anthropogenic and natural sources. Based on the previous surveys (Baumann et al., 2006; Mohankumar

et al., 2016; Sajedi-Hosseini et al., 2018; Twarakavi and Kaluarachchi, 2005; Uliasz-Misiak et al., 2022), eight environmental variables, including altitude, distance to roads, distance to waterbodies, geology, land use, rainfall, slope, and soil type were selected. In this study, all these factors were utilized as influencing factors to control the Ni contamination in groundwater. Maps of altitude and slope were extracted from the Aster Global Digital Elevation Model V003 satellite image. Distance to roads and distance to waterbodies were processed and estimated by the "Euclidean Distance" function in the ArcGIS environment from Thailand's national database of hydrology and transportation, respectively. Geology and soil type maps were provided by the Thailand Department of Mineral Resources. Land-use map was collected from Kanchanaburi Provincial Land Office. A rainfall map was created from Global satellite precipitation data (CHRS, http://chrs.web.uci.e.du). All influencing factors were prepared in raster format, with each pixel size of 30m x 30m.

3.2.3. Data about groundwater quality assessment

This study investigated previous publications relevant to groundwater quality assessment in reputable journals (Agrawal et al., 2021; Amiri et al., 2014; Asadi et al., 2019; Boateng et al., 2016; Chotpantarat and Thamrongsrisakul, 2021; Gulgundi and Shetty, 2018; Jha et al., 2020; Kawo and Karuppannan, 2018; Li et al., 2018; Rabeiy, 2018; Sadat-Noori et al., 2014; Sridharan and Senthil Nathan, 2017; Wali et al., 2019) (Supplementary Table S1). It was clear that the number of parameters was used differently in these studies. An investigation was then conducted statistics on the popularity of the parameters to select groundwater parameters with the use of above 50% for this study. Finally, potassium (K^+), sodium (Na^+), calcium (Ca^{2+}), magnesium (Mg^{2+}), chloride (Cl^-), sulfate (SO_4^{2-}), bicarbonate (HCO_3^-), nitrate (NO_3^-), pH, electric conductivity (EC), total dissolved solids (TDS) and total hardness (TH) were selected to assess the groundwater quality in Kanchanaburi, Thailand.

Next, I inherited groundwater samples of 116 wells in Nong Prue, Bo Phloi, Mueang Kanchanaburi, and Dam Makham Tia districts from the Groundwater Resource Department (DGR) conducted from January to April 2022 and then collected 64 groundwater samples in Lao Khwan, Huai Krachao, Phanom Thuan, Tha Maka, Tha Muang districts in May 2022 to assure the coverage rate (Fig. 3.12). A total of 180 groundwater samples were used in this study (Supplementary Table S2). All groundwater samples were contained in 500-ml polyethylene bottles and labeled corresponding to information of each well after getting from the groundwater wells. The measurements of EC and TDS were conducted by Hach 51800-10 sensION 5 Waterproof Conductivity Meter with an accuracy of ±0.5%. pH was measured by HQ40d Portable Multi Meter with an accuracy of ±0.01. Analysis of K^+ and Na^+ Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, HCO_3^-, NO_3^-, and TH were implemented at the DGR. All the physical parameters were measured in the field. The analysis results were checked with internal standards (Wisitthammasri et al., 2020) and with an ion charge balance error percentage of 0.3% (lower than 5%) (Li et al., 2018).

After getting information on groundwater samples from the DGR, the study conducted an analytical process by the SPPS software to determine the statistical values. In addition, the Piper Trilinear diagram was used to observe hydrogeochemical properties in groundwater using the Graph software (Ratchawang et al., 2022), and a correlation matrix was estimated using the "library(corrplot)" in the R environment to analyze the "Pearson" correlation between the groundwater parameters and the normalized EWQI (Cortadellas et al., 2017). Next, a geostatistical process was applied to develop the groundwater quality map. It was an important procedure because it could not collect all field data on the whole study area. In geostatistical analysis, the two typical interpolation methods are IDW and Kriging. The IDW is known as a deterministic method for unknown values based on the distance

weight between known values. Kriging is an interpolation technique that measures the unknown based on spatial statistics of known values (Jha et al., 2020). The IDW is used when data is a non-normal distribution. Meanwhile, Kriging is used when data is a normal distribution (Kerry and Oliver, 2007). For our case study, the normal distribution was determined by the Kolmogorov-Smirnov test. As a result, all data of groundwater parameters were non-normal distributions, so the IDW was applied to generate the groundwater parameter maps (Table 4.11).

3.3. Methodologies

3.3.1. Methodologies for mapping groundwater yield

3.3.1.1. Analytic Hierarchy Process (AHP)

AHP is a decision analysis technique based on the selection criteria of experts to evaluate a problem (Wang et al., 2009). Because of its effectiveness, it is applied to many fields of science and technology, including groundwater potential mapping (Arulbalaji et al., 2019; Das et al., 2019; Razandi et al., 2015). It provides quick support by examining the weight and condition factors of groundwater potential (Kumar and Krishna, 2018). The limitation of the AHP approach is the dependence on expert's knowledge and experience, which can lead to a few inconsistencies in some cases (Velasquez and Hester, 2013). In this study, five groundwater experts belonging to the DGR were asked to perform the AHP analysis. The experts were requested to provide ranking scores in the pairwise comparisons based on how they think a variable (influencing factor) or variable's subclass might be more important than the other for getting groundwater yield >10 m^3/h. Then, the influencing factors were integrated according to Saaty's 1–9 scale based on the experts' evaluation (Saaty, 1990). The influencing factors were independently determined such that only the odd scores were used for the AHP technique. These scores were defined based on different levels: 1 - equal importance, 3 - somewhat more important, 5 - much more

important, 7 - very much more important, and 9 - absolutely more important. A pairwise comparison matrix was utilized to collate the hierarchy of variables in the next step (Table 4.1). Eventually, the weight of factors on outcomes was calculated using a diagonal matrix (Table 4.2). Equations (2) and (3) were utilized to test the consistency ratio of the AHP model (Saaty, 1990):

$$CI = \frac{\lambda_{max} - n}{n - 1} \tag{2}$$

$$CR = \frac{CI}{RI} \tag{3}$$

where CI is the consistency index, λ_{max} is the highest eigenvalue in the comparison pairs, n is the number of factors ($n = 8$), CR is the consistency ratio, and RI is the random consistency index. The RI value was detailed in the previous study (Alonso and Lamata, 2006). The AHP is accepted when CR is less than 0.1 (Malczewski, 1999). In this model, the groundwater potential index was calculated based on the weight of the influencing factors and Saaty's 1–9 scale of subclasses, which was formulated as per equation (4) (Al-Djazouli et al., 2021; Arulbalaji et al., 2019; Owolabi et al., 2020):

$$GWPI_{AHP} = \sum W_i R_j, \tag{4}$$

where $GWPI_{AHP}$ is the groundwater potential index, W_i is the weight of the influencing factor i[th], and R_j is the Saaty's 1–9 scale of subclass j[th] of influencing factor i[th] assigned by experts. After achieving the $GWPI_{AHP}$, the study applied equation (5) to normalize this indicator into the scaling of 0 to 1.

$$X_{Normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{5}$$

where $X_{Normalized}$ is the normalized value, X is the original value, and X_{min} and X_{max} are the original minimum and maximum values.

3.2.1.2. Frequency Ratio (FR)

FR, a statistical algorithm, is used to calculate the probability of a particular object or phenomenon (Bonham-Carter, 1994). The FR model was employed to determine the possibility of groundwater occurrence with a particular yield in a basin in a previous study (Guru et al., 2017). Influencing factors were considered independent variables, while groundwater yield was regarded as a dependent variable (Razandi et al., 2015). The advantage of the FR technique is that the output weights of conditional factors are based on statistical data. The drawback of RF is its dependence on the ground truth data and method to classify subclasses of influencing factors (Wang et al., 2020). For our case study, the FR model was deployed based on the dataset of wells with a groundwater yield >10 m^3/h and the number of pixels of each variable's subclass to calculate the groundwater potential index. The equations for the FR approach were expressed as follows. Firstly, the FR value of each class was calculated by equation (6):

$$FR = \frac{(\frac{E}{TE})}{(\frac{F}{TF})} \tag{6}$$

where FR is the probability of occurrence of groundwater yield >10 m^3/h on each subclass of influencing factor, E is the number of well locations with groundwater yield >10m^3/h involving each subclass of influencing factor, TE is the total number of well locations, F is the number of pixels in each subclass of influencing factor, and TF is the total number of pixels of influencing factor. Secondly, the relative frequency (RF$_{FR}$) was calculated as follows:

$$RF_{FRi} = \frac{FR_{ij}}{\sum FR_j} \tag{7}$$

where RF$_{FRi}$ is the relative frequency of the ith subclass of jth influencing factors (j ranges from 1 to 8). FR_i is the frequency ratio value of ith subclass. $\sum FR_j$ is the total frequency ratio value of jth influencing factors. Next step, the predicted value (PV) was calculated by equation (8):

$$PVj = \frac{RF_{FRij\_max} - RF_{RFij\_min}}{(RF_{FRij\_max} - RF_{FRij\_min})average} \tag{8}$$

where RF_{FRij_max} is the maximum RF$_{FR}$ value of jth influencing factors, RF_{RFij_min} is the minimum RF$_{FR}$ value of jth influencing factors. The calculated values (RF & FV) were used to determine a set of many pixels with a groundwater potential index (GWPI) by equation (9):

$$GWPI_{FR\ model} = \sum(RF_{FRi} \times PV_j) \tag{9}$$

Similar to the AHP, the GWPI$_{FR\ model}$ was normalized into a scale of 0 to 1 using equation (5).

3.3.1.3. Random Forest (RF)

RF is an assembly algorithm for classification or regression based on multiple decision trees in a predictive model (Breiman, 2001). The tree's branches are formed from attributes of variables. The number of variables and decision trees must be determined before a model is implemented (Liaw and Wiener, 2002). In the GWP research, the RF model has been applied to define the potential location and reserves of groundwater based on well data and influencing factors (Naghibi et al., 2017a; Rahmati and Melesse, 2016; Zabihi et al., 2016). The training data for the RF models are encoded in a binary tree algorithm, which generates the nodes in a decision tree. The prediction error is estimated and influencing variables are permuted (Catani et al., 2013). However, the RF model overfits when training datasets are small or when there are too many input features (Wang et al., 2020). In our case study, an RF model was created to classify groundwater yield into two classes of yield below (class 0) and above (class 1) 10 m^3/h based on the eight input influencing factors. Because the groundwater yield data was imbalanced, stratified random sampling was used to create the training and testing data for the RF model building. The groundwater yield data for 1,601 wells was first split into two strata using the groundwater yield threshold of 10 m^3/h. The individual strata were then

randomly split into two subsets using a ratio of 70:30 before being combined into a training and testing set, respectively. This ensures that enough data from the two classes is available in both the training and testing sets for proper model training. The Caret package in the R program (Naghibi et al., 2017a) was used to train and tune the binary RF classifier. Ten-fold cross-validation was used to reduce overfitting on the training set during model training. Categorical input influencing factors, including geology, land use, and soil type, were converted into one-hot encoding (Nguyen et al., 2020a) depending on which method provides a better accuracy score. The cutoff threshold of prediction probability for class 1 (that is, groundwater yield >10 m^3/h) was set using the optimal value (optimal threshold) of the model in this study. The Gini importance (Kalantar et al., 2019) derived from the RF model training was used to rank the importance of the input variables (nominal weights) in classifying the groundwater yield. Regarding the calculation of the groundwater potential index, the study used the prediction probability from the RF model to generate the GWPZ map. Because probability below the optimal threshold was associated with groundwater yield of <10 m^3/h, the study used the optimal threshold value which was the minimum probability of the groundwater yield above the 10 m^3/h class. The adjusted probability dataset was then scaled between 0 and 1 using min-max normalization and used as the GWPI for the RF model.

3.3.1.4. Ensemble model

Ensemble refers to the combination of multiple alternative models to obtain better predictive performance than that obtained from individual ones (Opitz and Maclin, 1999). Since these alternative models are often trained using different algorithms or different datasets, they tend to cancel each other's weaknesses out and improve the overall accuracy (Muavhi et al., 2021; Rajaee et al., 2019; Yan et al., 2019).

In this study, to create an ensemble model of AHP, FR, and RF for GWP mapping, the study averaged the GWPIs from the three individual models. These GWPIs have been normalized to be on the same scale (0-1) in previous models.

3.3.1.5. Mapping of groundwater potential zones

For visualization purposes and better decision-making, the normalized GWPI was delineated for potential zones with groundwater yields above 10 m^3/h using Jenk's natural break function in the ArcGIS environment (Kumar et al., 2016). The potential zones were classified into three levels, high, moderate, and poor.

3.3.1.6. Validation for maps of groundwater potential zones

The validation of a model is essential for determining its predictive power (Chen et al., 2019c). The Receiver Operating Characteristic (ROC) curves and the Area Under the Curve (AUC) metrics (Naghibi et al., 2017a) were used to compare the performance of the three GWP models in this study. The ROC curves were obtained by plotting the cumulative area of GWPI zones represented by the validation wells on the y axis against (1 – GWPI) on the x axis (Fig. 3.14). The AUC, the area limited by the ROC space, demonstrates the accuracy of a prediction system by describing the system's ability to accurately predict the occurrence or non-occurrence of pre-defined "events" (Jothibasu and Anbazhagan, 2016). The AUC value ranges from 0 to 1. For our case study, a perfect GWP model should have all validation wells fall into zones with a GWPI of 1 (or a 1-GWPI value of 0), corresponding to the AUC of 1. This is because validation wells have a groundwater yield above 10 m^3/h, and therefore their probability, or "potential", is always 1. However, perfect GWP models are unlikely to exist. A good predictive GWP model aims to cover most of the validation wells in the high GWPI zones (that is, zones with low 1- GWPI). In other words, the closer the AUC value is to 1, the better the GWP model is.

To construct a ROC curve, the study classified the corresponding GWPI map

into 100 classes (incremented by 0.01 unit). Validation wells were also plotted on the classified map and the area of the GWPI zones that have the validation wells was recorded. The area of GWPI zones was normalized between zero and one by dividing its raw value by the total area represented by all validation wells. The GWPI and the corresponding zonal area represented by the validation wells were sorted in descending order of the GWPI. The cumulative area was then calculated by adding a value to the previous one in the sequence.
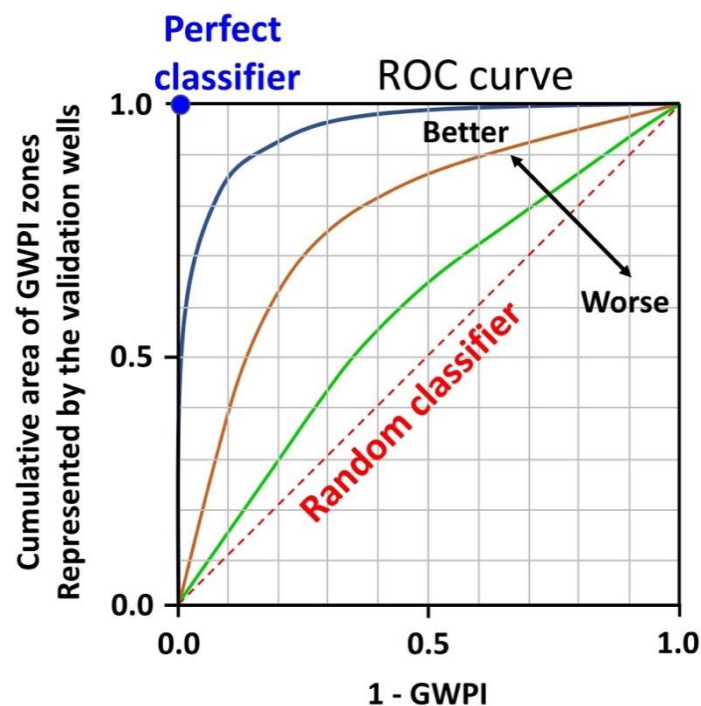
Fig. 3.14. Receiver Operating Characteristic curves and Area Under the Curve metrics

3.3.2. Methodologies for mapping Ni contamination risk zone in groundwater

The research process was carried out the following steps (Fig. 3.15): Step one is to build a database for the model. The input data includes the position of boreholes, information about influencing factors, and classification of Ni contamination and non-contamination. Groundwater samples are classified into contaminated and uncontaminated based on the Thailand groundwater quality standard for drinking purposes and were coded as 0 and 1. 0 is the uncontaminated

sample, and 1 is the contaminated well by Ni element. The database is divided into a ratio of 70:30 to train the model and test the results (126 samples for training; 54 samples for validating). Step two sets the weighted criteria and ranks the influencing factors based on contribution level and predicted value based on the Maxent technique, respectively. Step three calculates the AHP weight of influencing factors on the Ni contamination in groundwater using the AHP technique. Step 4 maps the Ni contamination risk zones in groundwater in the study area. Step 5 is validation.
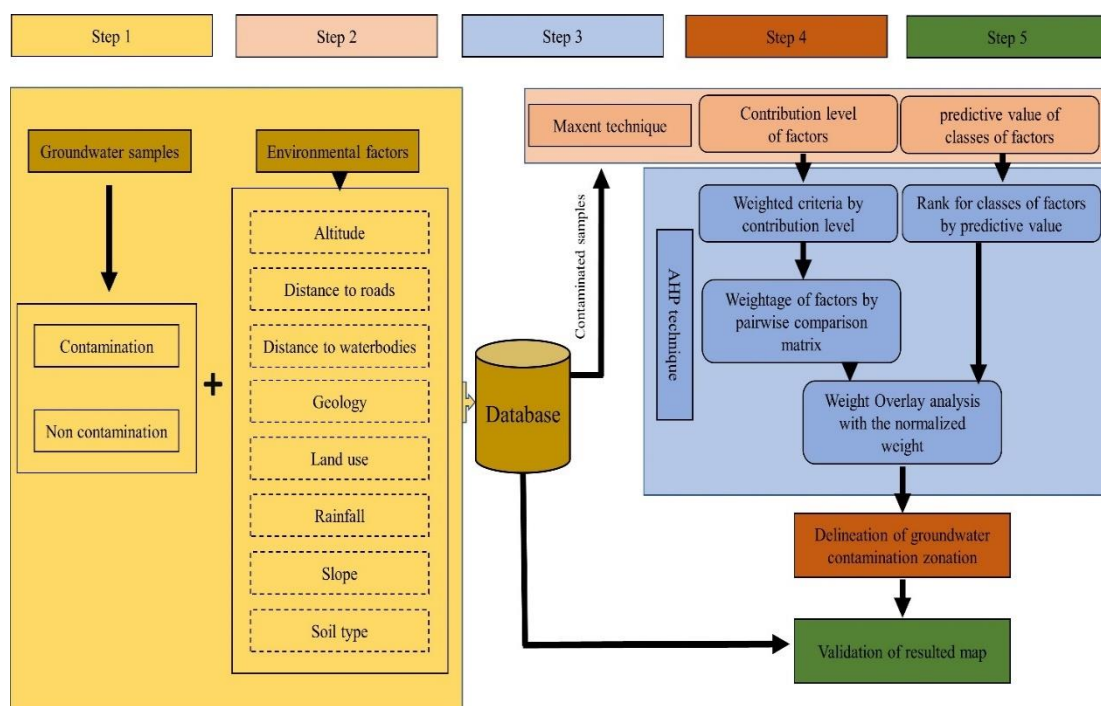
Fig 3.15. Flowchart for delineating the Ni contamination risk zones in groundwater

3.3.2.1. Maxent

Maxent is a statistical technique based on the maximal distribution of known outcomes (Mousazade et al., 2019). In this technique, the probability distribution function is built based on environment variables. The prediction result uses the probability distribution function to calculate the phenomenon's occurrence probability (Reddy and Dávalos, 2003). However, the maxent technique works towards a uniform distribution without considering the constraints of the environment variables (Kaky et al.,

2020). In this study, Maxent was used to estimate the maximum occurrence probability of heavy metal pollution in the study area and the contribution level of environmental factors. The Maxent mathematical formulas are as follows (Mousazade et al., 2019):

$$P_{ij} = \frac{b}{a} \tag{10}$$

$$(P_{ij}) = \frac{P_{ij}}{\sum_{i=1}^{sj} P_{ij}} \tag{11}$$

$$H_j = -\sum_{i=1}^{sj}(P_{ij})log_2(P_{ij}) \text{ ; j = 1; 2; 3; ...; n} \tag{12}$$

$$H_{jmax} = log_2 S_j \tag{13}$$

$$I_j = \frac{H_{jmax} - H_j}{H_{jmax}} \tag{14}$$

$$W_j = I_j P_{ij} \tag{15}$$

Where i is symbol for the ith influencing factors (i ranges from 1 to 8), j is symbol for the jth class of the influencing factors, a is the area of each class of each influencing factor, b is the number of contaminated groundwater samples in the jth class of the ith influencing factor, P_{ij} is the occurrence probability of Ni contamination in the jth class of the ith influencing factor, (P_{ij}) is the density probability, H is the entropy values, H_{jmax} is the maximum entropy value. The computing contribution level of influencing factors to the Ni contamination risk and the predicted value are performed in the R program environment with the SDM package (Algorithm 1).

Algorithm 1. Maxent approach

1: # Install package "SDM" and "Java environment"

2: # Import databases

3: # Construct maxent model

4: MX = max('environmental factors', 'train data')

5: # Getting the contribution variable

6: plot(MX)

7: # Get the predicted value

8: response(MX)

3.3.2.2. Analytic Hierarchy Process (AHP)

AHP is a technique in making decisions. In the AHP technique, the weighted criteria of the conditional variables are usually determined based on experts' opinions on with scale of 1 to 9 (Saaty, 1990). The odd scores are defined according to different levels: 1 - equal importance, 3 - somewhat more important, 5 - much more important, 7 - very much more important, 9 - absolutely more important. Meanwhile, the even scores are defined as intermediate values when compromise is needed (Saaty, 1990). In this study, the influencing factors were determined independently so that the odd scores were only used to set the weight criteria for the AHP model. The weighted criteria of influencing factors and the ranks of their classes were assigned based on the contribution level and the predicted values from the Maxent method, respectively. For the weighted criteria, the contribution level was normalized by a common scale based on Eq. (16), and then applied as following principle: 0-0.2 was assigned 1, 0.2-0.4 was assigned 3, 0.4-0.6 was assigned 5, 0.6-0.8 was assigned 7, 0.8-1 was assigned 9. For the ranks of classes, the scores were given from 1-9 (only odd scores) based on the predicted values. The higher the predicted value was, the higher the score was, meaning that the probability of contamination occurrence is high. After the weighted criteria were set, calculating the AHP weights of the influencing factor were performed in a pairwise comparison matrix. In the AHP approach, the consistency ratio less than 0.1 was used to evaluate the matrix results (Arulbalaji et al., 2019; Saaty, 1990). The consistency ratio was estimated according to the Eqs. (17) and (18):

$$X_{normalized} = \frac{CL - CL_{minimum}}{CL_{maximum} - CL_{minimum}} \qquad (16)$$

$$CI = \frac{\lambda_{max} - n}{n - 1} \qquad (17)$$

$$CR = \frac{CI}{RI} \qquad (18)$$

where $X_{normalized}$ is the normalized value of contribution level, CL is the value of contribution level of the ith influencing factors, CI is the consistency index, n is the number of environmental factors, λ_{max} is the highest eigenvalue of pairwise comparison matrix, CR is consistency ratio, RI is the random index of the number of environment variables. The RI value is detailed in the study of (Alonso and Lamata, 2006).

3.3.2.3. Calculating potential contamination index and mapping the nickel contamination risk zones

The potential contamination index is a dimensionless quantity. In this study, the potential contamination index was used to predict groundwater's Ni contamination risk. Based on the previous studies about potential prediction (Achu et al., 2020; Adiat et al., 2012; Arulbalaji et al., 2019; Dar et al., 2021; Saranya et al., 2020), the potential contamination index of Ni in groundwater for each location was estimated according to the following formula:

$$GWCPI = Al_wAl_f + DtR_wDtR_f + DtW_wDtW_f + Geo_wGeo_f + LU_wLU_f + R_wR_f + SL_wSL_f + S_wS_f \qquad (19)$$

Where GWPCI is the potential contamination index of Ni in groundwater, Al is the altitude, DtR is the distance to roads, DtW is the distance to waterbodies, Geo is the geology, LU is the land use, R is the rainfall, SL is the slope, and S is the soil, w is the weight of influencing factor, f is rank of the classes of the influencing factor.

The potential contamination index map of Ni in groundwater was based on the value of pixels calculated from equation (19). The GWCPI values were classified into three different levels, including high, moderate, low, using the Natural Breaks function in ArcGIS software.

3.3.2.4. Validation for contamination risk map of Ni in groundwater

The results verification is an essential step in scientific and forecasting studies (Barzegar et al., 2018). In the current study, the data for validating the output map was based on 54 groundwater samples (30% database for testing). The receiver operating characteristic (ROC) and statistical indicators were proposed to evaluate the model's performance and accuracy of the output map (Tien Bui et al., 2019). The ROC is a graphical plot of a curve that represents the relationship between sensitivity and 1 – specificity or true positive rate and false positive rate (Pham et al., 2019). The ROC value is the area under the curve, namely the area under curve (AUC) (Tien Bui et al., 2019). The AUC value changes from 0 to 1. AUC value close to 1 means the high model's performance, and vice versa. The statistical indicators include positive predictive value (PPV), negative predictive value (NPV), accuracy, sensitivity, specificity. These statistics are calculated based on the equation (20, 21,22,23,24) and confusion matrix (Table 3.1) (Pham et al., 2019):

$$Sensitivity = \frac{TP}{TP+FN} \tag{20}$$

$$Specificity = \frac{TN}{TN+FP} \tag{21}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{22}$$

$$PPV = \frac{TP}{TP+FP} \tag{23}$$

$$NPV = \frac{TN}{TN+FN} \tag{24}$$

Where TP is the amount of pixels correctly classified as contamination samples (positive results), TN is the amount of pixels correctly classified as "non-contamination samples" (negative results), FP is the amount of pixels incorrectly classified as "contamination samples" (positive results), FN is the amount of pixels incorrectly classified as "contamination samples" (negative results).

Table 3.1. Matrix for observed and predicted values

		Result	
		Polluted samples	Non-polluted samples
Observation	Polluted samples	TP	FN
	Non-polluted samples	FP	TN

3.3.3. Methodologies for mapping groundwater quality

3.3.3.1. EWQI computation

Entropy is a technique to calculate objective quality parameter weights (Peiyue et al., 2010). It overcomes the weighting system's subjectivity. The entropy creates a network of information to assess the indirect connections of variables. The influenced weight of a variable is mainly determined by the difference of alternative values when using another variable. Data is less information when the entropy value is high, and therefore the result is more unpredictable. In the procedures of calculating the EWQI, each parameter is given an entropy weight (Wu et al., 2018). In this study, groundwater samples and physicochemical parameters are assigned in the order of i from 1 to 180 (i.e., 180 groundwater samples) and j from 1 to 12 (i.e., twelve parameters), respectively. Mathematically, the formulas for computing the EWQI are presented as follows:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots a_{1j} \\ a_{21} & a_{22} & \dots a_{1j} \\ a_{i1} & a_{i2} & \dots a_{ij} \end{bmatrix} \tag{25}$$

$$b_{ij} = \frac{x_{ij} - x_{ij\,min}}{x_{ij\,max} - x_{ij\,min}} \tag{26}$$

$$B = \begin{bmatrix} b_{11} & b_{12} & ... b_{1j} \\ b_{21} & b_{22} & ... b_{2j} \\ b_{i1} & b_{i2} & ... b_{ij} \end{bmatrix} \tag{27}$$

$$P_{ij} = \frac{(1+b_{ij})}{\sum_{i=1}^{m}(1+b_{ij})} \tag{28}$$

$$e_j = -\frac{1}{\ln m}\sum_{i=1}^{m} P_{ij}\ln P_{ij} \tag{29}$$

$$w_j = \frac{(1-e_j)}{\sum_{i=1}^{m}(1-e_j)} \tag{30}$$

$$q_j = \frac{C_j}{S_j} \times 100 \tag{31}$$

$$EWQI = \sum_{j=1}^{n} w_j \times q_j \tag{32}$$

where A is the Eigenvalue matrix, b_{ij} is standardized value, B is a standard-grade matrix, P_{ij} is performance indices, e_j is the information entropy, w_j is the entropy weight, C_j is the concentration of parameters j, S_j is the permissible standard value, q_j is the quality rating scale. After achieving the EWQI, the study has applied equation (9) to normalize this indicator into the scaling of 0 to 1.

$$EWQI_{Normalized} = \frac{EWQI - EWQI_{min}}{EWQI_{max} - EWQI_{min}} \tag{33}$$

where $EWQI_{Normalized}$ is the normalized value, EWQI is the original value, $EWQI_{min}$ and $EWQI_{max}$ are the original minimum and maximum values. For visualization purposes and better decision making, the normalized EWQI was classified into five levels of groundwater quality, including excellent quality ($EWQI_{Normalized}$ of 0-0.2), good quality ($EWQI_{Normalized}$ of 0.2-0.4), moderate quality ($EWQI_{Normalized}$ of 0.4-0.6), poor quality ($EWQI_{Normalized}$ of 6-0.8), extremely poor quality ($EWQI_{Normalized}$ of 0.8-1).

3.3.3.2. Cross-validation and bootstrap techniques

In this study, the initial database, including 16 columns and 180 rows, was designed based on a matrix structure in which the columns were the information on the groundwater parameters and the rows were the information on groundwater

samples (Supplementary Table S2). In the initial database, groundwater parameters were the independent variables, the normalized EWQI was the target variable. Cross-validation and bootstrap techniques were then applied to build training and testing data subsets for the ML models. Cross-validation and bootstrap are resampling methods (Kohavi, 1995). While the cross-validation divides the initial dataset into multiple subsets for training and testing (Berrar, 2019), the bootstrap generates multiple subsets from the initial database after resampling with replacement (Chernick, 2012). For the cross-validation, the initial database was split into five subsets, each subset is called a data block (fold) comprised of 36 groundwater samples (Fig. 3.16a). For the bootstrap, the initial database was divided into 100 multiple subsets after resampling with replacement (Fig. 3.16b).
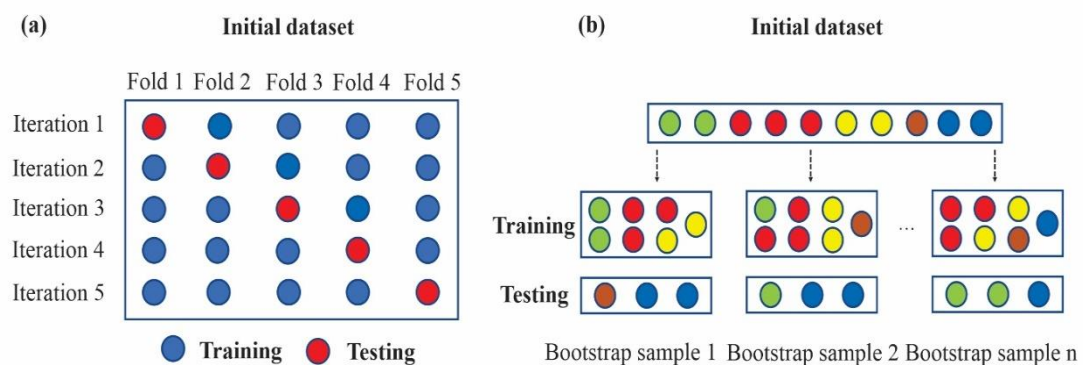
Fig. 3.16. Cross-validation (a) and bootstrap (b)

3.3.3.3. Random forest (RF)

Achievement in the application of the RF model has been recorded remarkably in branches of environmental science such as landslides (Cheng et al., 2021), floods (Schoppa et al., 2020), air pollution (Kumar, 2018), land use management (Wu et al., 2021), and water quality (Wang et al., 2021). The RF, an ensemble learning method, is established from the decision trees. In the RF model, a decision tree is an option for forecasting results, in which tree branches are a combination of data classes based on different datasets and attributes (Ahmad et al., 2017). An ensemble output of the RF is

an aggregate of the decision trees, therefore, the information on the decision trees complements each other, resulting in a model with low bias and low variance (Goldstein et al., 2011). As a corollary, the RF model solves overfitting and underfitting problems on the validation/test subset. An RF model conducts the task of classification or regression (Liaw and Wiener, 2002). In the RF algorithm, the number of decision trees (n-tree) and attribute information (m-try) must declare to determine the optimal model. However, the RF model returns results in a long time because it processes many predictions. For our case study, the RF algorithm was implemented in the R program environment using the "Caret" package. To start, the study built the control subjects using the "trainControl function" based on the cross-validation and bootstrap, respectively. Then, the study integrated the control subjects into the RF algorithm, called RF-CV and RF-B models, respectively. The role of RF in this study was to conduct a regression task because the normalized EWQI is a continuous variable ranging between 0 and 1. The value of the normalized EWQI for locations in the study area was conducted by the prediction function.

3.3.3.4. Artificial neural network (ANN)

The ANN, a machine learning model, is applied by many researchers to make predictions (Soltani Mohammadi et al., 2017; Tamiru and Dinka, 2021; Xu et al., 2021). The ANN is developed from a structure that is similar to the behavior of neurons and synapses in the human brain. It allows computer programs to recognize complex relationships in a dataset and solve the problem of prediction (Lee et al., 2017). The structure of an ANN model includes an input layer, hidden layers, and an output layer (Ahmad et al., 2017). The input layer represents input data, hidden layers represent the intermediate nodes that divide the input space into boundary regions, and the output layer represents the output of the neural network (Monteiro et al., 2021). Similar to the RF, the ANN can perform classification or regression missions. The ANN works based on the correlation of input data to make decisions. However,

ANN only works with numeric data and requires large datasets, which requires the categorical data need to be generalized coding (Tayyebi and Pijanowski, 2014). For our case study, the groundwater parameters were designated as the input layer, and the normalized EWQI was appointed as the output layer. The "nearalnet" package was performed directly in the R program. The cross-validation and bootstrap techniques were integrated into the ANN algorithm to build the ANN models, called the ANN-CV and ANN-B, respectively. The default number of hidden layers was one with a threshold of 0.05 because the more hidden layer is, the more complex the model is (Bedi et al., 2020). The ANN-CV and ANN-B algorithms were presented in supplementary materials.

3.3.3.5. Model validation

Modeling can lead to deviations between forecast and observed values. Therefore, model validation is an essential step for any model. For our case study, coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) were utilized to accurately validate and compare the performance between the models (Bedi et al., 2020). The R^2 is an indicator to explain the coefficient of variance explanation, which describes the relationship between a set of parameters on the target variable. The value of R^2 varies from zero to one. The R^2 goes to one, the model is good, and vice visa (Santhi et al., 2001). The RMSE, an error-index statistic, is used to compare the performance of models. The lower the RMSE is, the more reliable model is. Similar to RMSE, MAE is also applied to evaluate the models. The value of MAE goes to 0, which means that model is more suitable (Chai and Draxler, 2014).

$$R^2 = \frac{\sum_{i=1}^{n}(EWQI_n - \overline{EWQI_n}) * (EWQI_p - \overline{EWQI_p})}{(\sum_{i=1}^{n}(EWQI_n - \overline{EWQI_n})^2)^{0.5} * (\sum_{i=1}^{n}(EWQI_n - EWQI_p)^2)^{0.5}} \tag{34}$$

$$MAE = \frac{\sum_{i=1}^{n}(EWQI_n - EWQI_p)}{n} \tag{35}$$

$$RMSE = \sqrt{\frac{1}{n}((EWQI_n - EWQI_p)^2} \tag{36}$$

where $EWQI_n$ is the estimated value of normalized EWQI, $EWQI_p$ is the predicted value of the normalized EWQI.

Chapter 4. RESULTS AND DISCUSSION

4.1. Exploring spatial distribution of groundwater yield in Kanchanaburi, Thailand using AHP, FR, and RF

4.1.1. Influencing factors on groundwater yield

Altitude: Altitude affects groundwater distribution as water often moves down the elevation gradient. Therefore, it is an indicator of groundwater potential (Chen et al., 2019c). In the study area, altitude values ranged from 2 m to 1,812 m and were divided into five subclasses: 2–50 (2,331 km^2), 50–100 (2,278 km^2); 100–150 (1,263 km^2); 150–200 (2,150 km^2); and >200m (11,368 km^2) (Fig. 4.1a).

Distance to faults: Faults are hydraulic barriers hindering horizontal groundwater movement. Therefore, it is an important factor to invest groundwater yield (Naghibi and Pourghasemi, 2015). The distance to faults map of Kanchanaburi was generated with five subclasses: 0–10,000, 10,000–20,000, 20,000–30,000, 30,000–40,000, and >40,000 m, in turn accounting for an area of 14,384, 2,480, 1,601, 857, and 67 km^2, respectively (Fig. 4.1b).

Distance to waterbodies: Distance to waterbodies affects the infiltration of groundwater and the recharge capacity of an aquifer (Chen et al., 2019c). The distance to waterbodies map was classified into five subclasses: 0–2,000 (13,315 km^2), 2,000–4,000 (3,956 km^2), 4,000–6,000 (1,388 km^2), 6,000–8,000 (489 km^2), and >8,000 m (242 km^2) (Fig. 4.1c).

Slope: Slope determines excess rainfall and runoff accumulation and thus influences the residence time of water on the surface and the ability of water to seep into the underground (Nguyen et al., 2020c). The slope in the study area was

classified into five subclasses: 0°–3° (3,304 km^2), 3°–6° (3,820 km^2), 6°–10° (3,122 km^2), 10°–16° (3,070 km^2), and >16° (6,073 km^2) (Fig. 4.1d).

Geology: Geology is important for the distribution of groundwater. Indeed, good aquifers are often found on fractured rocks (limestone, basalts, conglomerates, sandstone) and unconsolidated sediments (sand, gravel), while poor aquifers are composed of igneous rocks (granite, shale) or sediments (mud, clay, slit) (MacDonald and Davies, 2000). In the study area, there are fourteen geological units, including Carboniferous-Permian (CPk), Devonian (D), Cambrian (E), Cambrian-Ordovician (EO), Jurassic (J), Cretaceous (Kgr), Ordovician (O), Pre-Cambrian (PE), Permian (Pr), Quaternary (Q), Silurian-Devonian (SD), Silurian-Devonian-Carboniferous (SDC), Tertiary (Tmm), Triassic (Tr) in which Q covers the highest area (5,474 km^2), followed by SDC (3,180 km^2), O (2,647 km^2), Pr (2,430 km^2), Tr (1,472 km^2). Moreover, other geological units occupy a small area of Kanchanaburi (Fig. 4.1e).

Land use: Land use reveals the current state of the surface coating of a geographic feature, which has a close relationship with groundwater through residence time and the infiltration process of surface water. It represents soil moisture, surface roughness, movement and recharge capacity of groundwater (Sameen et al., 2019). In Kanchanaburi, the land use factor comprises five land cover units, reclassified into agricultural land (5,323 km^2), miscellaneous land (454 km^2), forest land (12,224 km^2), urban and built-up land (557 km^2), and waterbody (506 km^2) (Fig. 4.1f). Forest is primarily distributed in the West; moreover, agricultural, miscellaneous, urban and built-up lands concentrate in the East.

Rainfall: Rainfall greatly influences the recharge of groundwater through surface input and subsurface infiltration (Khoshtinat et al., 2019). The rainfall map was divided into five subclasses as follows: <1,500 (8,091 km^2), 1,500–1,650 (4,223 km^2), 1,650–1,800 (5,051 km^2), 1,800–1,950 (1,964 km^2), and >1,950 mm/year (60 km^2)

(Fig. 4.1g).

Soil type: Soil controls the infiltration and retention rate of water, which is decided by soil structure and texture (Martínez-Santos and Renard, 2020). In Kanchanaburi, there are fourteen soil types in which clayey sand is the dominant type, comprising ~11,189 km^2 of the total area, followed by sandy loam (2,759 km^2) and clay loam (1,987 km^2). The others cover only a small part of the area and are sparsely distributed in the province (Fig. 4.1h).
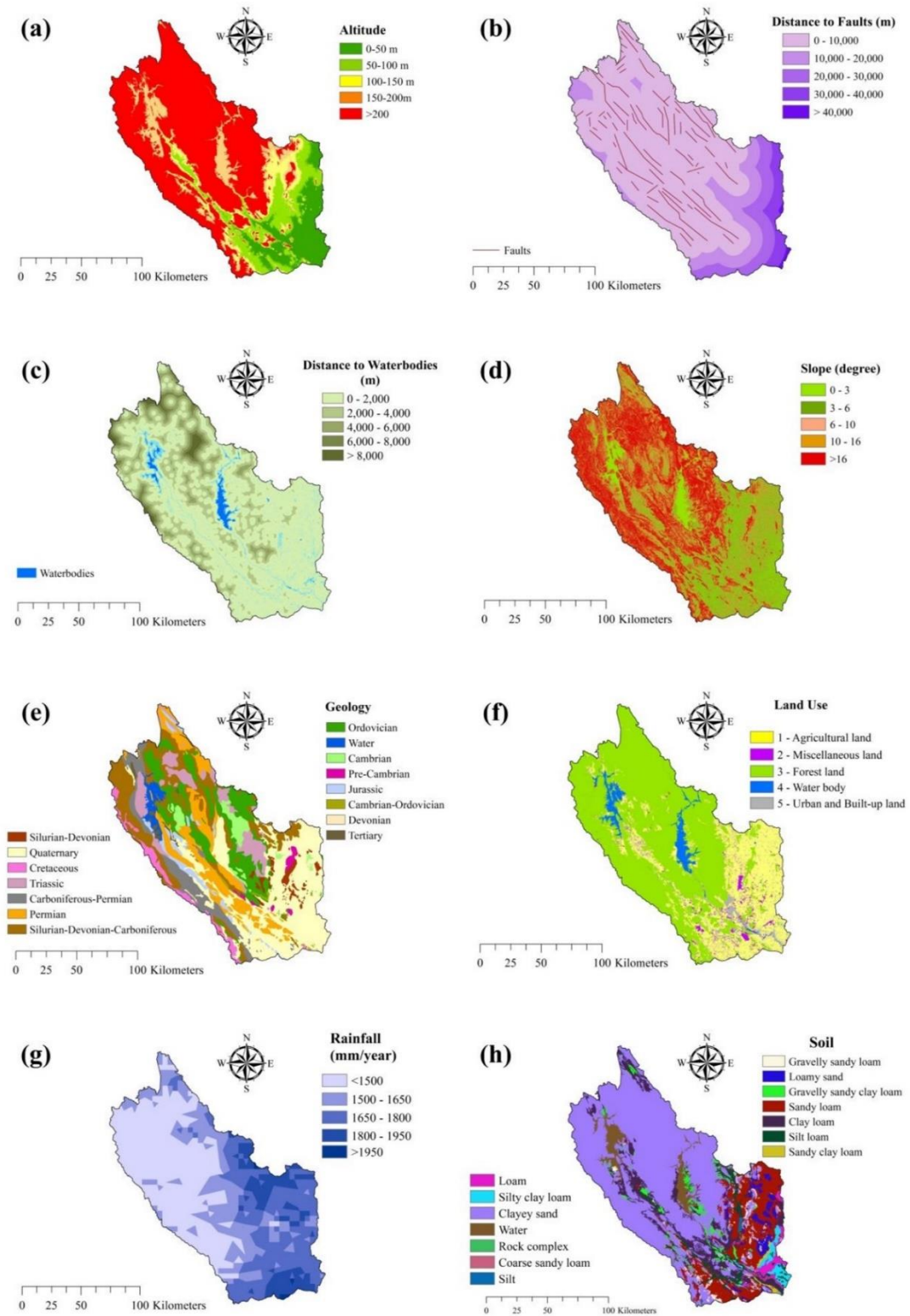
Fig. 4.1. Influencing factors include **(a)** altitude, **(b)** distance to faults, **(c)** distance to waterbodies, **(d)** slope, **(e)** geology, **(f)** land use, **(g)** rainfall, **(h)** soil type

4.1.2. Application of AHP for delineating groundwater potential zones

The AHP approach provides the weight of the influencing factors for the presence of wells with a groundwater yield above 10 m³/h. A pairwise comparison matrix of influencing factors determined by the judgment of experts in the groundwater field was used to calculate their diagonal matrix (Tables 4.1 and 4.2). The highest weight of 28.66% was found in altitude and distance to faults. It was followed by distance to water and slope at 14.40%. Meanwhile, the figures for geology, rainfall, and soil type were only 3.95%, and land use was the least impact factor with a weight of 2.03% (Tables 4.2 and 4.33). The consistency ratio was 0.055 (Table 4.2) revealing the reasonability of the pair-wise comparison process.

Table 4.1. Pairwise comparison matrix of influencing factors

AHP	Al	DF	DW	Slope	Geology	LU	Rainfall	Soil type	Weight criteria
Al	1	1	3	3	7	9	7	7	9
DF	1	1	3	3	7	9	7	7	9
DW	1/3	1/3	1	1	5	7	5	5	7
Slope	1/3	1	1	1	5	7	5	5	7
Geology	1/9	1/7	1/7	1/3	1	3	1	1	3
LU	1/7	1/5	1/5	1	3	1	1/3	1/3	1
Rainfall	1/7	1/5	1/5	1	3	1	1	1	3
Soil type	1/9	1/7	1/7	1/3	1	1/3	1/3	1	3
Geometric Mean	3.21	3.21	8.74	8.74	27.33	42.00	27.33	27.33	

Abbreviations: Al – Altitude; DF – Distance to faults; DW – Distance to waterbodies; LU – Land use

While the high groundwater potential zones were observed mainly in the East

of Kanchanaburi Province, the poor groundwater potential zones were found in the western area (Fig. 4.3a). Specifically, 59.88% and 22.23% of GWP were represented by poor and moderate potential, respectively, while high potential accounted for 17.89% (Fig. 4.4).

Table 4.2. Diagonal matrix of influencing factors

AHP	Al	DF	DW	Slope	Geology	LU	Rainfall	S	Sum	Weight (Wi)	Eigenvalue (λ)
Al	0.31	0.31	0.34	0.34	0.26	0.21	0.26	0.26	2.29	28.66	8.55
DF	0.31	0.31	0.34	0.34	0.26	0.21	0.26	0.26	2.29	28.66	8.55
DW	0.10	0.10	0.11	0.11	0.18	0.17	0.18	0.18	1.15	14.40	8.43
Slope	0.10	0.10	0.11	0.11	0.18	0.17	0.18	0.18	1.15	14.40	8.43
Geology	0.04	0.04	0.02	0.02	0.04	0.07	0.04	0.04	0.32	3.95	8.07
LU	0.03	0.03	0.02	0.02	0.01	0.02	0.01	0.01	0.16	2.03	8.11
Rainfall	0.04	0.04	0.02	0.02	0.04	0.07	0.04	0.04	0.32	3.95	8.07
S	0.04	0.04	0.02	0.02	0.04	0.07	0.04	0.04	0.32	3.95	8.07

Λ max = 8.55; CI = 0.078; RI = 1.4056; CR = 0.055

Abbreviations: Al – Altitude; DF – Distance to faults; DW – Distance to waterbodies; LU – Land use, S – Soil type

Table 4.3. Normalized weights and Saaty's scale of factors' subclasses in the AHP model

Influencing factors	Subclasses	Weight (Wi)	Saaty's scale (Rj)
Altitude	2-50 m	28.66	9
	50-100 m		7
	100-150 m		5
	150-200 m		3
	>200 m		1
Distance to faults	0 to 10,000 m	28.66	1
	10,000 to 20,000 m		3
	20,000 to 30,000 m		5
	30,000 to 40,000 m		7
	> 40,000 m		9
Distance to waterbodies	0 to 2,000 m	14.40	9
	2,000 to 4,000 m		7
	4,000 to 6,000 m		5
	6,000 to 8,000 m		3
	>8,000 m		1
Slope	0° to 3°	14.40	9
	3° to 6°		7
	6° to 10°		5

	10° to 16°		3
	>16°		1
Geology	Silurian-Devonian		7
	Quaternary		9
	Cretaceous		1
	Triassic		1
	Carboniferous-Permian		3
	Permian		5
	Silurian-Devonian-Carboniferous		3
	Ordovician	3.95	3
	Water		9
	Cambrian		7
	Pre-Cambrian		1
	Jurassic		3
	Cambrian-Ordovician		1
	Devonian		1
	Tertiary		1
Land use	Agricultural land		7
	Miscellaneous land		3
	Forest land	2.03	5
	Waterbody		9
	Urban and Built-up land		1
Rainfall	<1,500 mm/year		1
	1,500 to 1,650 mm/year		3
	1,650 to 1,800 mm/year	3.95	5
	1,800 to 1,950 mm/year		7
	>1,950 mm/year		9
Soil type	Gravelly sandy loam		7
	Loamy sand		9
	Gravelly sandy clay loam		5
	Sandy loam		5
	Clay loam		3
	Silt loam		1
	Sandy clay loam	3.95	3
	Loam		5
	Silty clay loam		1
	Clayey sand		3
	Water		9
	Rock complex		3
	Coarse sandy loam		3
	Silt		3

4.1.3. Application of FR for delineating groundwater potential zones

The occurrence probability of groundwater well locations with a yield >10 m^3/h on influencing factors and its spatial relationship with influencing factors was observed using the FR approach (Table 4.4, Fig. 4.3b). This spatial relationship was shown by the predicted value, which was considered an indicator for calculating the GWPI. The spatial relationship will be positive, average, and negative when the predicted values are >1, 1, and <1, respectively (Lee and Pradhan, 2006). It is clear that the well location with a yield of >10 m^3/h and certain influencing factors, including distance to faults, distance to waterbodies, land use, and rainfall, had a positive spatial relationship. Meanwhile, a negative spatial relationship was observed between well location and altitude, slope, geology, and soil type. Besides, the relative frequency is also another important indicator of the impact of an influencing factor's subclasses on the occurrence of groundwater yield above 10 m^3/h (Muavhi et al., 2021). In the influencing factor of altitude, the subclasses with a low value had a high relative frequency and vice versa. The relative frequency value was higher when the distance was far from the faults. The influencing factors showed a sharp contrast with distance from waterbodies, slope, and rainfall. About geological units, the analysis of the FR approach indicated that the Quaternary had the highest value of relative frequency (0.41), followed by the Cambrian (0.21), and Silurian-Devonian (0.17). According to land-use type, urban and built-up land had the highest value of relative frequency (0.90) followed by agricultural land (0.07). Also, an assessment of soil type revealed that the largest relative frequency values of 0.28 and 0.20 were found in silt and silt loam, respectively. The majority of areas were affected by moderate (43.58%) and poor (30.92%) potential, while the remaining areas were correlated with high potential (25.50%) (Fig. 4.4).

Table 4.4. Frequency ratio values of the influencing factors

Influencing factors	Subclasses	No. of pixels	% of pixels	No. of wells	% of Well	FR	RF$_{FR}$	PVj
Altitude	2-50 m	2,590,578	12.023	94	32.53	2.71	0.48	
	50-100 m	2,531,187	11.748	40	13.84	1.18	0.21	
	100-150 m	1,404,065	6.517	27	9.34	1.43	0.25	0.87
	150-200 m	2,389,108	11.088	9	3.11	0.28	0.05	
	>200 m	12,631,163	58.624	5	1.73	0.03	0.01	
Distance to faults	0 to 10,000 m	15,983,312	74.182	40	13.84	1.87	0.01	
	10,000 to 20,000 m	2,755,716	12.790	33	11.42	8.93	0.05	
	20,000 to 30,000 m	1,779,129	8.257	57	19.72	23.89	0.14	1.15
	30,000 to 40,000 m	952,586	4.421	34	11.76	26.61	0.16	
	>40,000 m	75,358	0.350	11	3.81	108.83	0.64	
Distance to waterbodies	0 to 2,000 m	14,795,142	68.667	173	59.86	0.87	0.93	
	2,000 to 4,000 m	4,396,017	20.403	1	0.35	0.02	0.02	
	4,000 to 6,000 m	1,542,558	7.159	1	0.35	0.05	0.05	1.70
	6,000 to 8,000 m	453,408	2.104	0	0.00	0.00	0.00	
	> 8,000 m	268,976	1.248	0	0.00	0.00	0.00	
Slope	0° to 3°	3,671,525	17.040	67	23.18	1.36	0.40	
	3° to 6°	4,244,957	19.702	68	23.53	1.19	0.35	
	6° to 10°	3,469,374	16.102	33	11.42	0.71	0.21	0.72
	10° to 16°	3,411,916	15.835	6	2.08	0.13	0.04	
	>16°	6,748,329	31.320	1	0.35	0.01	0.00	
Geology	Silurian-Devonian	402,951	1.870	2	0.69	0.37	0.07	
	Quaternary	6,082,732	28.234	152	52.60	1.86	0.36	
	Cretaceous	625,202	2.902	0	0.00	0.00	0.00	
	Triassic	1,636,212	7.595	0	0.00	0.00	0.00	
	Carboniferous-Permian	1,412,578	6.557	1	0.35	0.05	0.01	
	Permian	2,700,326	12.534	5	1.73	0.14	0.03	0.66
	Silurian-Devonian-Carboniferous	3,533,408	16.401	3	1.04	0.06	0.01	
	Ordovician	2,941,752	13.655	2	0.69	0.05	0.01	
	Water	412,851	1.916	0	0.00	0.00	0.00	
	Cambrian	679,377	3.153	5	1.73	0.55	0.11	
	Pre-Cambrian	183,734	0.853	2	0.69	0.81	0.16	
	Jurassic	638,309	2.963	1	0.35	0.12	0.02	

	Cambrian-Ordovician	105,767	0.491	1	0.35	0.70	0.14	
	Devonian	12,972	0.060	0	0.00	0.00	0.00	
	Tertiary	175,578	0.815	1	0.35	0.42	0.08	
Land use	Agricultural land	5,915,278	27.457	66	22.84	0.83	0.06	
	Miscellaneous land	504,868	2.343	5	1.73	0.74	0.05	
	Forest land	13,582,255	63.046	2	0.69	0.01	0.00	1.62
	Waterbody	921,919	4.279	0	0.00	0.00	0.00	
	Urban and Built-up land	619,084	2.874	102	35.29	12.28	0.89	
Rainfall	<1500 mm/year	8,990,313	41.727	6	2.08	0.05	0.01	
	1500 to 1650 mm/year	4,692,909	21.781	20	6.92	0.32	0.05	
	1650 to 1800 mm/year	5,612,822	26.051	84	29.07	1.12	0.19	1.12
	1800 to 1950 mm/year	2,182,294	10.129	63	21.80	2.15	0.37	
	>1950 mm/year	67,445	0.313	2	0.69	2.21	0.38	
Soil	Gravelly sandy loam	349,197	1.621	3	1.04	0.64	0.04	
	Loamy sand	514,108	2.387	7	2.42	1.01	0.06	
	Gravelly sandy clay loam	576,966	2.679	4	1.38	0.52	0.03	
	Sandy loam	3,066,194	14.235	68	23.53	1.65	0.09	
	Clay loam	2,207,956	10.250	21	7.27	0.71	0.04	
	Silt loam	559,393	2.597	20	6.92	2.66	0.15	
	Sandy clay loam	98,007	0.455	1	0.35	0.76	0.04	0.60
	Loam	372,193	1.728	29	10.03	5.81	0.33	
	Silty clay loam	366,614	1.702	20	6.92	4.07	0.23	
	Clayey sand	12,433,143	57.720	2	0.69	0.01	0.00	
	Water	921,425	4.278	0	0.00	0.00	0.00	
	Rock complex	37,538	0.174	0	0.00	0.00	0.00	
	Coarse sandy loam	2,031	0.009	0	0.00	0.00	0.00	
	Silt	35,696	0.166	0	0.00	0.00	0.00	

Abbreviations: FR - Frequency ratio; RF_{Fri} - Relative frequency of ith subclass; PVj - Predicted value of jth influencing factors

4.1.4. Application of RF for delineating groundwater potential zones

As previously suggested in section 2.3.3, the RF model was conducted by the binary regression algorithm. The importance of influencing factors was also observed

by the RF model (Fig. 4.2). The contribution level to groundwater well occurrences with a yield >10 m³/h of distance to waterbodies was the highest (22.00%), followed by the altitude factor (20.40%). Distance to faults, slope and soil type was the next in the contribution level with 16.87%, 15.90%, and 11.88%, respectively. The other factors were insignificant or had a low impact. The final GWP map from the RF approach is presented in Fig. 4.3c. The majority of regions had poor potential at 60.67% of the area, followed by the proportion of moderate potential (20.79 %). Meanwhile, only 18.54% of the area was defined by high GWP (Fig. 4.4).
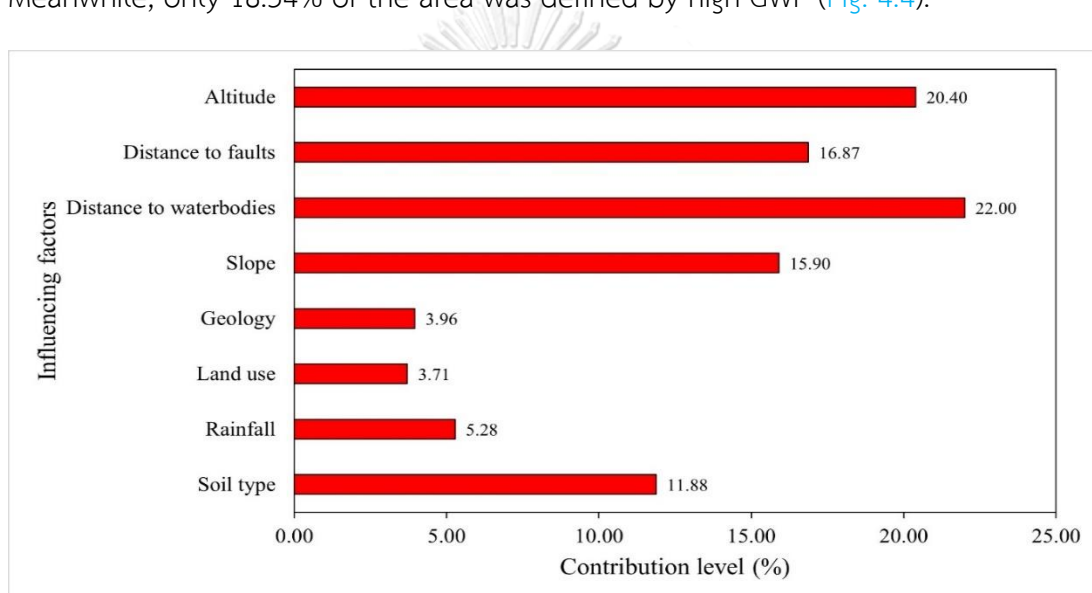
Fig. 4.2. Contribution level of influencing factors on groundwater yield >10 m³/h conducted by the RF

4.1.5. Application ensemble model of AHP, FR, RF for delineating groundwater potential zones

An ensemble model was created by averaging the normalized GWPIs predicted in this study by the AHP, FR, and RF models (Fig. 4.3d). Poor and moderate potential made up 36.42% and 30.15% of the area, respectively, while 33.43% of the area showed high potential (Fig. 4.4).
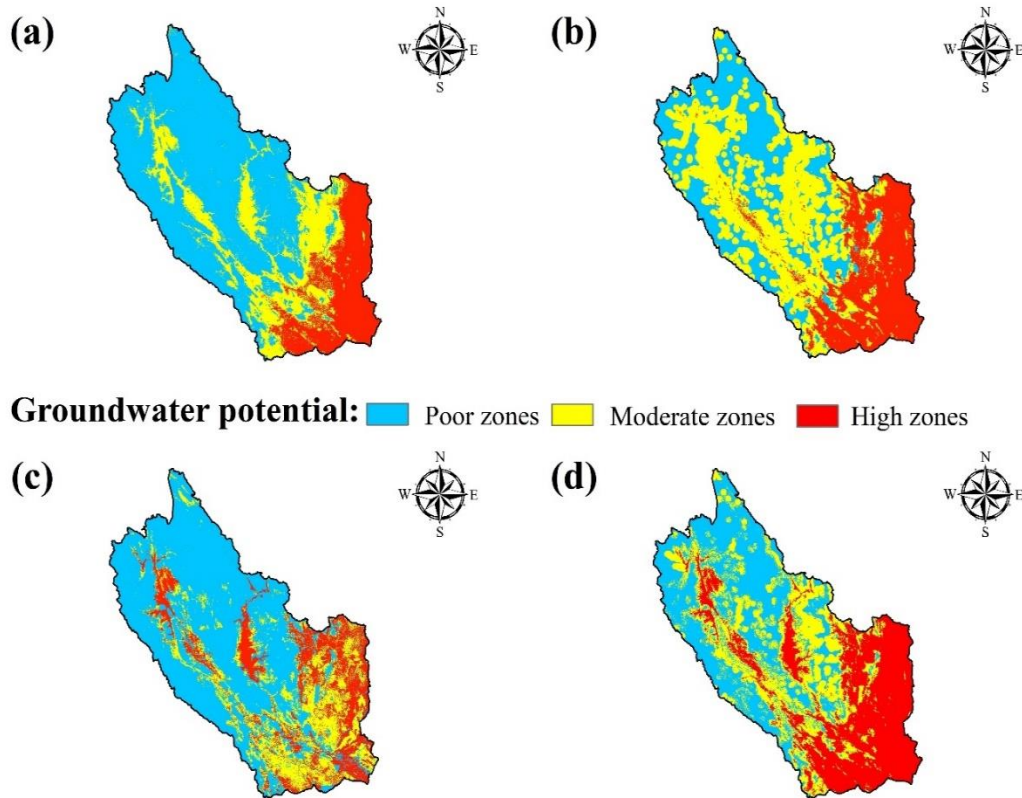
Fig. 4.3. Groundwater potential maps for yield >10 m³/h in Kanchanaburi Province, Thailand: **(a)** AHP model; **(b)** FR model; **(c)** RF model; **(d)** Ensemble model
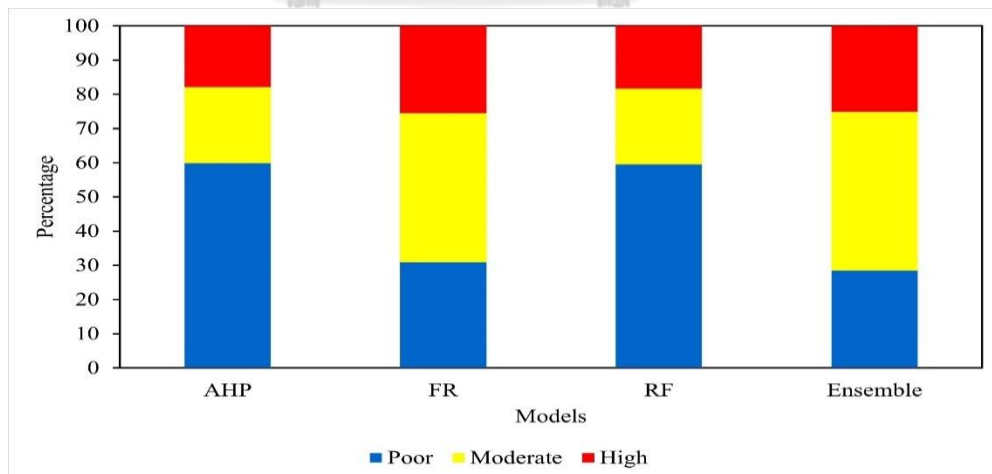
Fig 4.4. Area of GWP zones for groundwater yield >10 m³/h in Kanchanaburi Province

4.1.6. Assessment of the model's accuracy

The models' results were assessed by comparing the cumulative GPWIs of

the testing set via AUC value. The AUCs of the AHP, FR, RF, and ensemble models were 0.72, 0.74, 0.76, and 0.80, respectively (Fig. 4.5), revealing that all models achieved similarly good predictions. Nevertheless, individual models exhibited different behaviors. Specifically, the RF model did better at identifying the target groundwater yield at the GWPI above 0.6 (or 1-GWPI below 0.4) but worse at lower GWPI, compared to other models. The predictive power of FR was lower than that of the random classifier at the GWPI above 0.9 (that is, 1-GWPI below 0.1), but outweighed that of RF and AHP at the GWPI below 0.6. The AHP model, on the other hand, expressed monotonic behavior across the GWPI range. Combining the results from all models into an ensemble did improve the overall model prediction. The ensemble model achieved an AUC score of 0.8 and performed better for the GWPI values between 0.2 and 0.8, whereas all the individual models fell short.
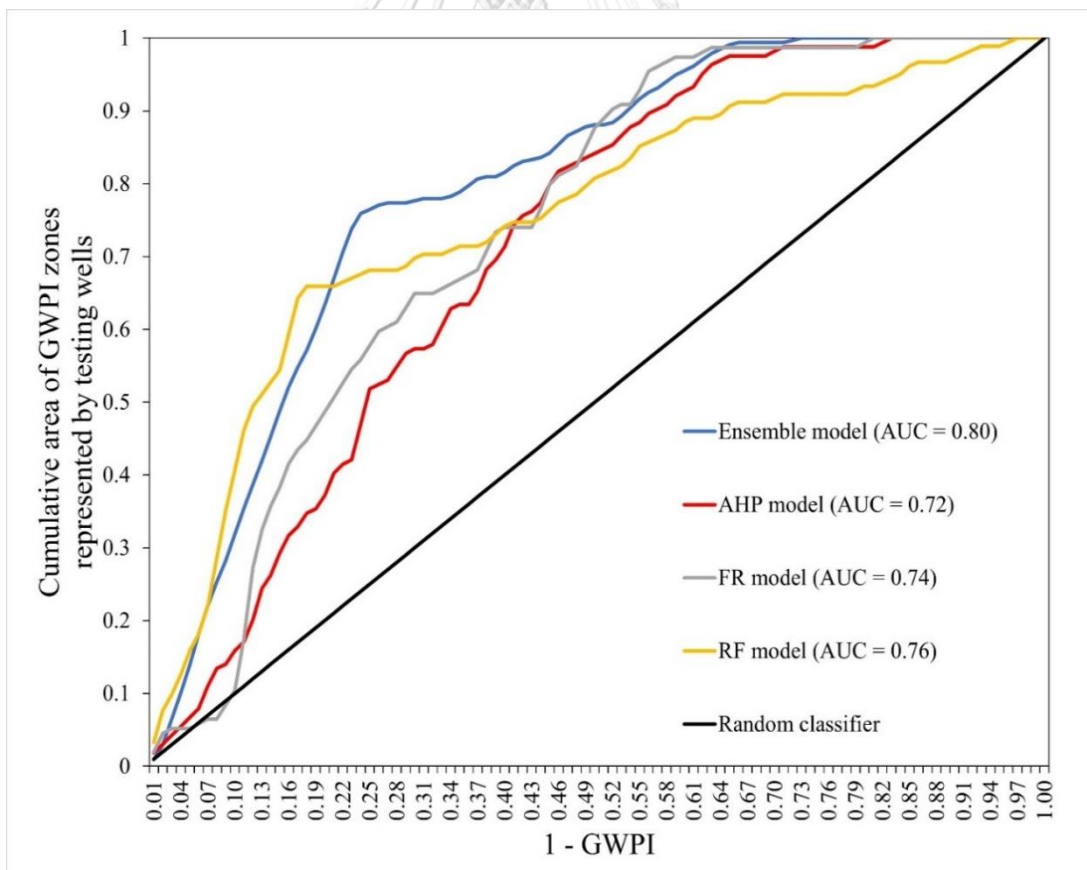
Fig. 4.5. Validation of models.

4.1.7. Discussion

Quantifying the importance of the influence factors on groundwater processes helps policy-makers improve water resource planning. In our study, the models used have different mechanisms to weigh the influencing factors as they belong to different approaches. The AHP model used a decision system approach, so the weights of influencing factors were assigned completely based on expert opinions and domain knowledge, while the FR and RF models assigned weights based on data analysis. The FR method eliminated human interference and simply assigned weights based on the number of wells with a groundwater yield exceeding 10 m^3/h that fell into different sub-classes of the influencing factors. The RF, on the other hand, used more sophisticated optimization algorithms to compute the weight for each influencing factor based on the information gained between all models that included a specific influencing factor and all models that excluded that factor. Our results showed that the AHP and RF methods had similar rankings for the top four most important factors, including altitude, distance to faults, distance to waterbodies, and slope, that influence the occurrence of wells with groundwater yields exceeding 10 m^3/h in the study area (Table 4.5). The FR identified distance to waterbodies and distance to faults as the first and third most important factors but assigned less weight to altitude and slope. This was because the FR employed a less sophisticated statistical approach, and thus might fail to capture the effects of interaction between the influencing factors. Nevertheless, our analysis indicated that the AHP could be a cost-effective and accurate method to identify the most important influencing factors for policy-making priorities.

Table 4.5. Ranking of influencing factors by a method based on the factor weights. The ranking score ranges from one to eight in descending order of importance

Influencing factor	AHP	FR	RF
Altitude	1	5	2
Distance to faults	2	3	3
Distance to waterbodies	3	1	1
Slope	4	6	4
Geology	5	7	7
Land use	6	2	8
Rainfall	7	4	6
Soil	8	8	5

Our analysis confirmed findings from previous studies regarding the influence of geophysical characteristics on groundwater yield. In particular, the study found that high groundwater yield was mostly observed in areas with an altitude below 100 m, which asserted the impact of gravity on groundwater. A similar study by Al-Abadi (2017a) also showed the occurrence of high groundwater yield in areas with an altitude below 99 m while Khoshtinat et al. (2019) indicated that the lower the altitude, the higher the GWP for their study region. Our study revealed the greater influence of faults as a natural barrier to impeding the occurrence of high groundwater yield, which was in line with a study by (Anderson and Bakker, 2008). However, some previous studies (Arulbalaji et al., 2019; Arunbose et al., 2021; Falah and Zeinivand, 2019) insisted that high fault density could be an advantage for the rechargeability of groundwater, which was often observed in areas with dominant vertical or mountainous areas with basement rocks (Chuma et al., 2013). The distance to waterbodies such as rivers, lakes, dams, and springs influenced groundwater yield because surface water was a recharging source of aquifers. The shorter

the distance to waterbodies was, the higher the GWP was. The relationship has also been discussed intensively in previous studies (Chen et al., 2018; Machiwal et al., 2011; Mosavi et al., 2020; Tahmassebipoor et al., 2016). The slope was also identified as one of the top influencing factors for high groundwater yield in our study with higher yields observed in areas with a slope of below 10°. Similar conclusions were drawn by other studies (Al-Abadi, 2017a; Ganapuram et al., 2009; Jothibasu and Anbazhagan, 2016; Kumar and Krishna, 2018).

The GWP with a yield >10 m^3/h of the lowland areas was higher than that of the mountainous areas in Kanchanaburi, where the terrain was flat with a low slope and altitude (Ahmed and Sajjad, 2018; Falah and Zeinivand, 2019; Shao et al., 2020) (Fig. 4.3). This finding was consistent with a previous hydrogeology survey conducted by the DGR in Kanchanaburi province (DGR, 2006). The GWP maps using the RF and ensemble models were considerably similar compared to the AHP and RF, respectively. This can be explained by the Jenks natural break function allowing the equivalent GWPI values to be grouped into a cluster to minimize the variance between classes in the mapping (Papaioannou et al., 2015). The poor GWP zones estimated by the AHP and RF models were 60% of the entire area, which is almost twice as higher as that computed by FR and ensemble. Meanwhile, the moderate GWP represented approximately 30% and 60% of the maps produced using AHP or RF and FR or ensemble models, respectively. The area with high GWP was in a range of 17.89 - 33% from model to model (Fig 4.4). The model selection for GWPZ mapping is an important but difficult decision as it often involves trade-offs among resources, time, expertise, and accuracy. These trade-offs are different for different study areas and objectives. For this study, the study showed that all models achieved similarly good prediction outcomes (Fig. 4.5). However, the RF and FR were more complex, costly, and time-consuming than the AHP model. Therefore, one might be better off just using the AHP model for similar regions where the ground truth data for groundwater yield is not

available. This will save time and costs as well as reduce the complexity. However, in other cases when the accuracy of the GWPZ mapping is a priority, the study recommend the use of complex machine learning models (e.g., RF and artificial neural networks) and/or an ensemble of different models and approaches to improve the overall predictive power, as illustrated by our analysis.

4.2. Exploring nickel contamination risk zones in groundwater in the eastern part of Kanchanaburi province, Thailand using Maxent and AHP

Based on the results from the delineation of groundwater yield potential zones (section 4.1), sections 4.2 and 4.3 focus on in the eastern part of Kanchanaburi province, Thailand, including Nong Prue, Lao Khwan, Bo Phloi, Huai Krachao, Phanom Thuan, Mueang Kanchanaburi, Tha Maka, Tha Muang, Dan Makham Tia districts, where groundwater yield was evaluated high and moderate potential.

4.2.1. Influencing factors on Ni contamination in groundwater

Eight influencing factors, including altitude, distance to road, distance to waterbodies, geology, land use, rainfall, slope, and soil type, were used to delineate the Ni contamination risk zones. Figure 4a-h shows the details of the influencing factors.

Altitude factor represents the influence of topography on groundwater depth and flow rate (Condon and Maxwell, 2015). When the flow rate and groundwater depth are significant, groundwater's chemical and physical characteristics can be altered in space and time (Alizamir et al., 2019). Altitude in the study area can be observed with elevation values ranging from 2 to 1812 meters above sea level, classified into five classes: 2–50 (2,265.64 km^2), 50–100 (1,819.74 km^2); 100–150 (881.59 km^2); 150–200 (578.12 km^2); and >200m (1,278.38 km^2) (Fig. 4.6a). The majority of the high elevation zones are on the northwest side.

Distance to roads is considered an essential factor in assessing groundwater

quality (Wang et al., 2018). Road transport activities contribute significantly harmful substances to groundwater through the dry and wet deposition of effluents, dust, and exhaust gases (Uliasz-Misiak et al., 2022). For our case study, distance to roads map was categorized into five classes: 0–2,000 (5,588.13 km^2), 2,000–4,000 (954.11 km^2), 4,000–6,000 (190.33 km^2), 6,000–8,000 (71.61 km^2), and >8,000 m (20.54 km^2) (Fig. 4.6b).

Waterbodies represent the close relationship between surface water and groundwater in the hydrological cycle. Changes in groundwater's chemical and physical composition are also influenced by the recharge and discharge processes (Lipczynska-Kochany, 2018). Distance to waterbodies factor was categorized into five classes: 0–2,000 (5,081.24 km^2), 2,000–4,000 (1,255.16 km^2), 4,000–6,000 (254.47 km^2), 6,000–8,000 (134.44 km^2), and >8,000 m (99.42 km^2) (Fig. 4.6c).

Geology plays a significant role in determining the chemical composition of groundwater because groundwater is contained in bedrock aquifer or host sediment (Golkarian and Rahmati, 2018). Twelve geological units are observed in the study area, including Tertiary (Tmm), Devonian-Carboniferous (SDC), Silurian-Devonian (SD), Quaternary (Q), Permian (Pr), Pre-Cambrian (PE), Ordovician (O), Cretaceous (Kgr), Jurassic (J), Cambrian-Ordovician (EO), Cambrian (E), Carboniferous-Permian (CPk), in which Q covers the highest area (4,413.86 km^2), followed by SDC (474.25 km^2), O (421.22 km^2), Pr (330.14 km^2), Tr (285.74 km^2) (Fig. 4.6d).

Land use reflects human activities on the land surface (Zhang et al., 2019). The wastes from human activities affect the quality of groundwater (McLay et al., 2001). There are five main land use types in the study area, including urban land (446 km^2), agricultural land (4,253 km^2), miscellaneous (362 km^2), forest land (1688 km^2), and waterbody (76 km^2) (Fig. 4.6e).

Rainfall is known as a recharge source of groundwater, meaning its chemical

components in the water can affect groundwater quality (Ali et al., 2020). In this study, the rainfall data was divided into five different regions as follows: <1,500 (149.73 km^2), 1,500–1,650 (1,692.15 km^2), 1,650–1,800 (3,859.18 km^2), 1,800–1,950 (1,062.97 km^2), and >1,950 mm/year (60.7 km^2) (Fig. 4.6f).

Slope reflects the direction and the speed of surface runoff, influencing the weathering and surface corrosion processes. Therefore, slope also affects surface water's and groundwater's chemical and physical composition (Arulbalaji et al., 2019). Slope value varies between 0° and 70° in the study area, classified into five classes: 0°–3° (1,904.08 km^2), 3°–6° (2,340.29 km^2), 6°–10° (1,238.63 km^2), 10°–16° (575.26 km^2), and >16° (765.22 km^2) (Fig. 4.6g). Regions with high slopes are concentrated in the western hilly areas.

Soil type is known as a natural filter of the earth's crust for groundwater, meaning chemicals can be partially trapped as groundwater seeps into the soil (Keesstra et al., 2012). Fourteen different texture types are found in the study areas, including gravelly sandy loam (314.28 km^2), loamy sand (462.70 km^2), gravelly sandy clay loam (519.27 km^2), sandy loam (2,759.57 km^2), clay loam (1,987.16 km^2), silt loam (503.45 km^2), sandy clay loam (88.21 km^2), loam (334.97 km^2), silty clay loam (329.95 km^2), clayey sand (11,189.83 km^2), water (829.28 km^2), rock complex (33.78 km^2), coarse sandy loam (1.83 km^2), silt (32.13 km^2) (Fig. 4.6h).
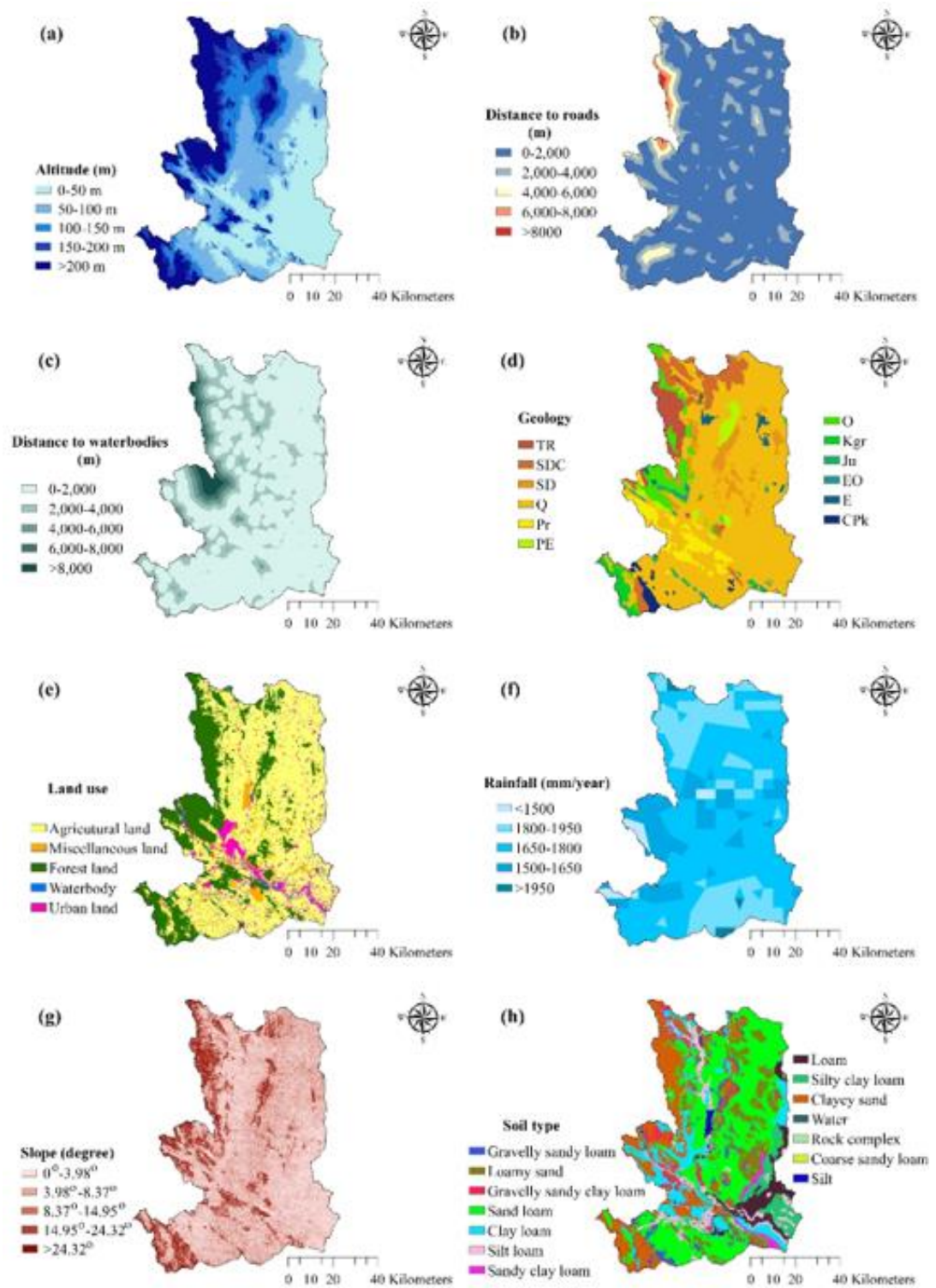
Fig. 4.6. Influencing factors: **(a)** altitude, **(b)** distance to roads, **(c)** distance to waterbodies, **(d)** geology, **(e)** land use, **(f)** rainfall, **(g)** slope, **(h)** soil type

4.2.2. Application of Maxent technique to estimate the contribution level and predicted value of influencing factors over Ni contamination.

The Maxent approach provides information about the contribution level of influencing factors to the occurrence ability of Ni contamination in groundwater and the predicted value of classes of influencing factors (Wahyudi et al., 2013). Fig. 4.7 illustrates the contribution level of influencing factors to Ni contamination determined by the Maxent technique. The result shows land use has the highest contribution of 65.4% to the delineation of Ni contamination in groundwater, followed by soil type (14.7%), slope (7.7%), and altitude (4.8%). The other factors contribute to the model by approximately 2%. Additionally, the impact of influencing factors on Ni contamination in groundwater is displayed in Fig. 4.8. The high response of Ni contamination in groundwater occurred at altitude of 0-50 m, distance to roads of 0-2000 m, distance to waterbodies of >8000 m, geology unit of Pre-Cambrian, land use of urban land, slope of 0°-3°, soil type of clay loam.
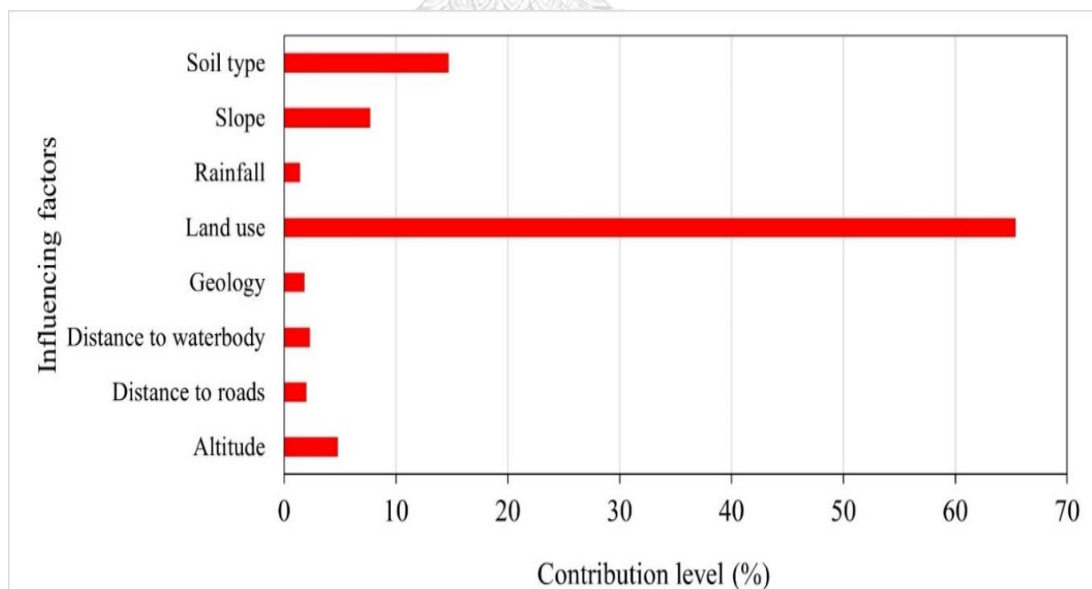
Fig. 4.7. Role of influencing factors in predicting Ni contamination in groundwater
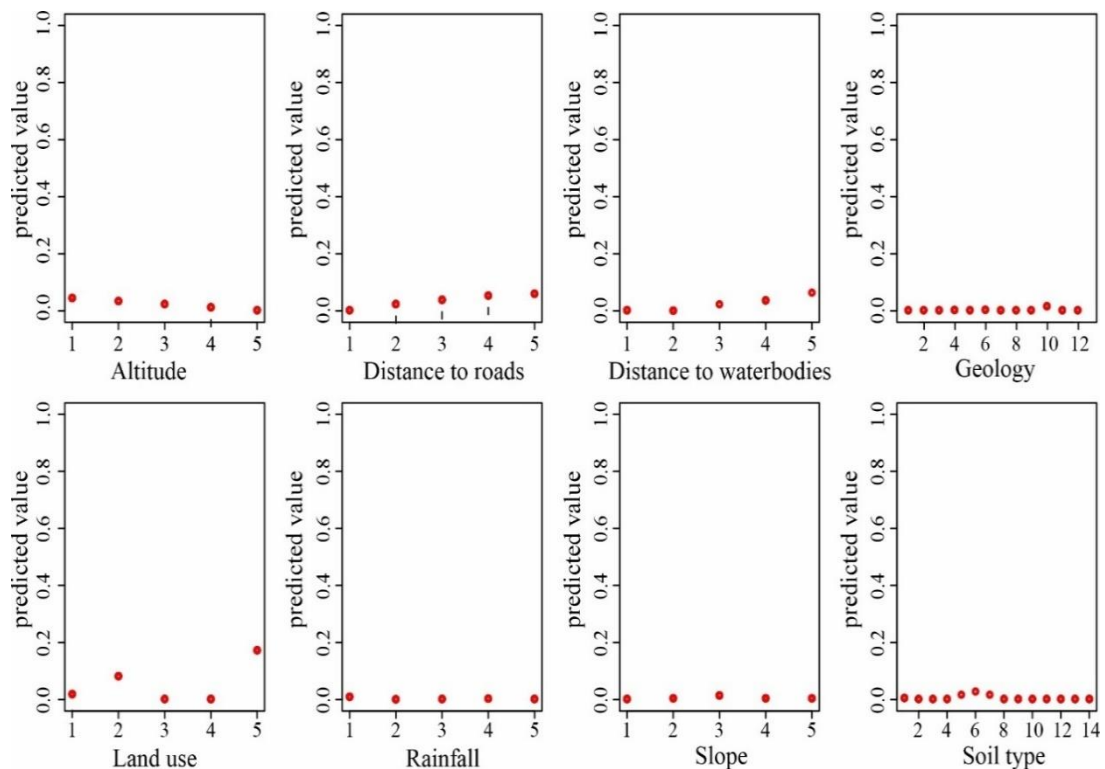
Fig. 4.8. The predicted values for classes of influencing factors on the occurrence ability of Ni contamination in groundwater

4.2.3. Application of AHP for calculating the weight of influencing factors

A pairwise comparison matrix was established from the contribution level of influencing factors to the occurrence ability of Ni contamination in groundwater (Table 4.6). Table 4.7 presents the weighted criteria of influencing factors for Ni contamination prediction in groundwater. The rank of classes was set based on the predictive value (Table 4.8). The result of the AHP analysis indicates that the consistency ratio of the matrix was 0.049. This consistency ratio proves that the analytical result of the model was acceptable (Arulbalaji et al., 2019; Saaty, 1990). Among influencing factors, land use had the highest AHP weight at 48.711, followed by soil type and geology. The other influencing factors had the same AHP weight of 4.287.

Table 4.6. Pairwise comparison matrix of influencing factors

Factors	Altitude	DtR	DtW	Geology	Land use	Rainfall	Slope	Soil type	Criteria
Altitude	1	1	1	1	9	1	1	3	1
DtR	1	1	1	1	9	1	1	3	1
DtW	1	1	1	1	9	1	1	3	1
Geology	1	1	1	1	9	1	1	3	1
Land use	1/9	1/9	1/9	1/9	1	1/9	1/9	1/7	9
Rainfall	1	1	1	1	9	1	1	3	1
Slope	1	1	1	1	9	1	1	3	1
Soil type	1/3	1/3	1/3	1/3	7	1/3	1/3	1	3

Table 4.7. Matrix for calculating the consistency ratio

Factors	Altitude	DtR	DtW	Geology	Land use	Rainfall	Slope	Soil type	Weighted sum	Geometric mean	AHP weight
Altitude	0.056	0.056	0.056	0.056	0.061	0.056	0.056	0.033	0.428	0.054	5.351
DtR	0.056	0.056	0.056	0.056	0.061	0.056	0.056	0.033	0.428	0.054	5.351
DtW	0.056	0.056	0.056	0.056	0.061	0.056	0.056	0.033	0.428	0.054	5.351
Geology	0.056	0.056	0.056	0.056	0.061	0.056	0.056	0.033	0.428	0.054	5.351
Land use	0.500	0.500	0.500	0.500	0.553	0.500	0.500	0.700	4.253	0.532	53.158
Rainfall	0.056	0.056	0.056	0.056	0.061	0.056	0.056	0.033	0.428	0.054	5.351
Slope	0.056	0.056	0.056	0.056	0.061	0.056	0.056	0.033	0.428	0.054	5.351
Soil type	0.167	0.167	0.167	0.167	0.079	0.167	0.167	0.100	1.179	0.147	14.737

λ_{max} = 8.376; RI_8 = 1.4056; CI = 0.038 (<0.1)

Table 4.8. Normalized weights and rank of factors' classes

Factors	Normalized Weight	Code	Classes	Rank
Altitude	5.351	1	0-50 m	9
		2	50-100 m	7
		3	100-150 m	5
		4	150-200 m	3
		5	> 200m	1
Distance to roads	5.351	1	0-2000 m	9
		2	2000-4000 m	7
		3	4000-6000 m	5
		4	6000-8000 m	3
		5	>8000 m	1
Distance to waterbodies	5.351	1	0-2000 m	1
		2	2000-4000 m	3
		3	4000-6000 m	5

		4	6000-8000 m	7
		5	>8000 m	9
Geology	5.351	1	Silurian-Devonian	1
		2	Quaternary	1
		3	Cretaceous	1
		4	Triassic	1
		5	Carboniferous-Permian	1
		6	Permian	1
		7	Silurian-Devonian-Carboniferous	1
		8	Ordovician	1
		9	Cambrian	1
		10	Pre-Cambrian	3
		11	Jurassic	1
		12	Cambrian-Ordovician	1
Land use	53.158	1	Agricultural land	3
		2	Miscellaneous land	5
		3	Forest land	1
		4	Water bodies	1
		5	Urban land	9
Rainfall	5.351	1	<1500 mm/year	1
		2	1500 to 1650 mm/year	1
		3	1650 to 1800 mm/year	1
		4	1800 to 1950 mm/year	1
		5	>1950 mm/year	1
Slope	5.351	1	0°-3°	9
		2	3°-6°	7
		3	6°-10°	5
		4	10°-16°	3
		5	>16°	1
Soil type	14.737	1	Gravelly sandy loam	1
		2	Loamy sand	1
		3	Gravelly sandy clay loam	1
		4	Sandy loam	5
		5	Clay loam	5
		6	Silt loam	5
		7	Sandy clay loam	3
		8	Loam	3
		9	Silty clay loam	3
		10	Clayey sand	1
		11	Water	1
		12	Rock complex	1
		13	Coarse sandy loam	1
		14	Silt	1

4.2.4. Map of Ni contamination risk zone in groundwater

Groundwater is known as a drinking water source in many regions of the world. Using groundwater with high Ni concentrations for domestic purposes will lead to cancer risk and health complications (Egbueri, 2020). Thus, a map of groundwater's Ni contamination risk zones is necessary in any region of the world. In this study, the map of groundwater's Ni contamination risk zone was identified by the potential contamination index to delineate the Ni contamination risk zone in groundwater. The output map was spatially categorized as low, moderate, and high zones (Fig. 4.9). Overall, 24.79% of the area (1691.82 km^2) was very low contamination risk of Ni, whereas the zone of high Ni contamination risk accounted for around 6.56% (447.65 km^2). Moderate contamination risk zone of Ni occupied 68.65% of the study area (Table 4.9).
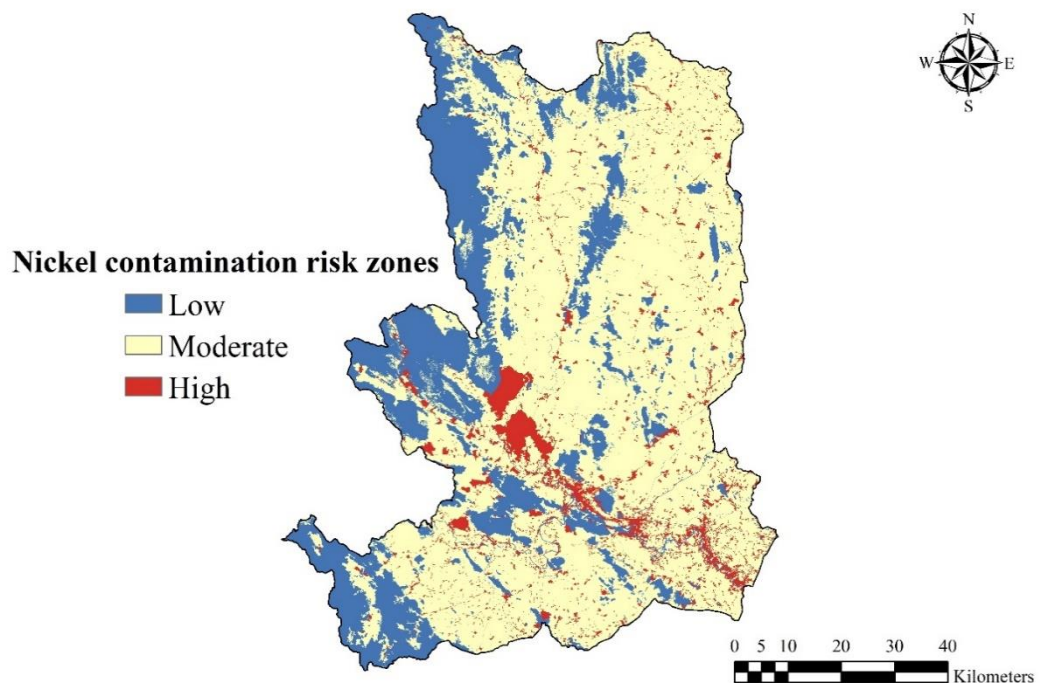
Fig. 4.9. Map of Ni contamination risk zones in groundwater

Table 4.9. Statistics of Ni contamination risk zones in groundwater in the study area

Level	Number of pixels	Area (km²)	Percentage (%)
Low risk	1879835	1691.85	24.79
Moderate risk	5204419	4683.98	68.65
High risk	497389	447.65	6.56

4.2.5. Validation

The ROC technique and statistical indicators were utilized to confirm the Ni contamination risk zones in groundwater. A total of 54 groundwater samples (30% of the database) were used to validate the model's performance validation and the accuracy of the output map. The validation results clearly show that the model's AUC value index was 0.86 (Fig. 4.10), and the accuracy of the resulting map was 85%. The results of other statistical indicators are presented in Table 4.10. From the model validation result and the accuracy of the output map, it can be concluded that combining the maxent method to set the weighted criteria and the AHP technique to delineate the heavy metal contamination risk zone in groundwater is an effective model.
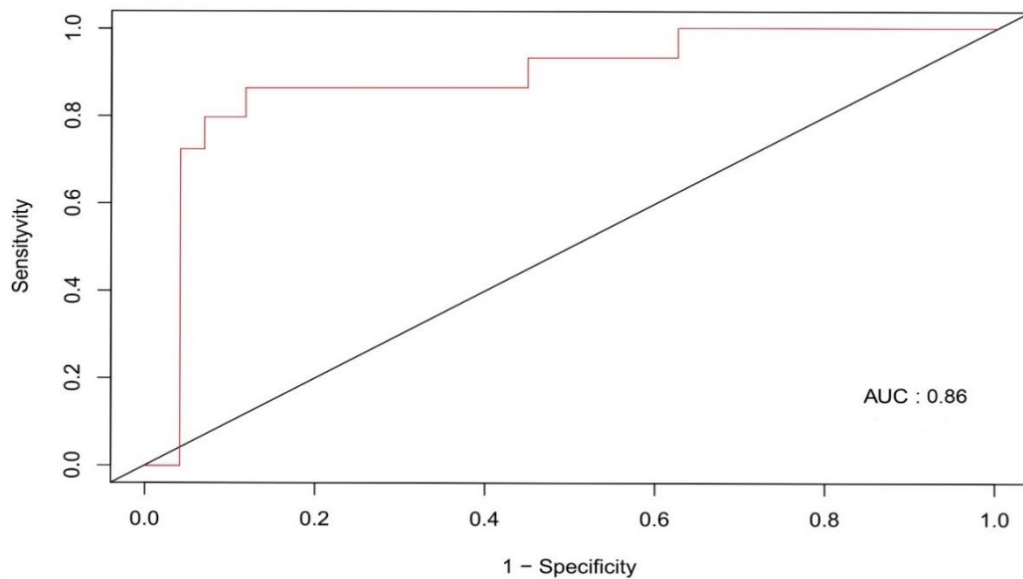
Fig. 4.10. Model performance

Table 4.10. Results of model's validation and output map

TP	TN	FP	FN	PPV	NPV	Sensitivity	Specificity	Accuracy
6	40	4	4	0.60	0.91	0.60	0.91	0.85

4.2.6. Discussion

Mapping the Ni contamination risk zone in groundwater is necessary for the long-term management of water resources. In this study, the maxent algorithm was selected to replace the expert's opinions in the AHP technique. By comparison to the previous studies (Arunbose et al., 2021; Azimi et al., 2018; Chakraborty et al., 2016), the criteria score has been given based on the contribution level of influencing factor to Ni contamination in groundwater from the output of maxent approach, which help to eliminate the subjective from judges as using the AHP technique. Additionally, the limitation of controlling factor was also broken compared to previous studies on groundwater vulnerability and risk (Barzegar et al., 2018). In our case study, eight influencing factors was considered conditional variables for predicting Ni contamination risk, including altitude, distance to roads, distance to waterbodies,

geology, land use, rainfall, slope, and soil type, which had never done in assessing groundwater contamination risk.

Regarding sensitivity analysis, the model shows the role of influencing factors in determining the Ni contamination, in which land use is the influencing factor with the largest contribution, followed by soil type. The analysis results indicated that land use and soil type factors explain approximately 67.8% variation in Ni contamination in groundwater. Our analysis unveiled the correlation between influencing factors and Ni contamination risk. The higher the altitude value is, the lower the Ni contamination is. Distance to roads and slopes are inversely proportional to the Ni contamination risk, while distance to water bodies is proportional to the Ni contamination risk. Geologically, the Ni contamination risk in groundwater of the Pre-Cambrian unit is higher than that of other units. Regarding land use, the urban areas were defined as the highest-risk zone with Ni contamination in groundwater, followed by the miscellaneous zones, agricultural zones, and forest land. Indeed, human activities were the leading cause of Ni contamination in groundwater (Egbueri, 2020; Khatri and Tyagi, 2015). In terms of soil type, the Ni contamination in groundwater of locations with clay loam, sandy loam, and clay loam is higher than that of locations with other soil textures. The findings of the present study were consistent with previous studies on the causes of groundwater contamination (Egbueri, 2020; Sajedi-Hosseini et al., 2018; Singha et al., 2020; Tiankao and Chotpantarat, 2018). In addition, our study indicated that the effect of rainfall is the same in any value.

Compared to other administrative districts, the groundwater in the Mueang Kanchanaburi district was found to be a high Ni contamination risk, raising concern over groundwater pollution management, particularly in urban areas. Therefore, strict policies for discharging waste materials should be applied in this region (Hu et al., 2014). The Ni contamination in groundwater used in the present study is considered

a striking example to show the geographical distribution of possibly polluted regions and to investigate the critical cause in identifying contaminated risk areas using the Maxent and AHP techniques. The AUC value indicated that the used model was a relatively effective solution for mapping Ni contamination susceptibility in groundwater.

4.3. Deciphering groundwater quality in the eastern part of Kanchanaburi province, Thailand using RF and ANN

4.3.1. Groundwater quality in Kanchanabui

Table 4.11 provides information on twelve physicochemical parameters in groundwater samples in Kanchanaburi, Thailand. The mean value of Ca^{2+} (97.09 mg/l) and HCO_3^- (461.94 mg/l) in the groundwater of the study area were the above-recommended limits. The Kolmogorov-Smirnov test shows the p-value of physicochemical parameters ranging from 0 to 0.002. In terms of calculating the EWQI, the entropy weight of NO_3^- was the highest, followed by SO_4^{2-}, Cl^-, K^+, Na^+, EC, Mg^{2+}, Ca^{2+}, TH, TDS, HCO_3^-, and pH. Groundwater parameters exceeded the allowed limit in some groundwater samples except for the Cl^-. The concentration of HCO_3^-, Ca^{2+}, and Mg^{2+} were normally found above the allowed limit with a number of 173, 132, and 70 groundwater samples, respectively. Regarding hydrogeochemical properties, alkaline earth (Mg+Ca) dominated over alkalis (Na+K), while, and weak acids (HCO_3) and strong acids (SO_4+Cl) were not clearly distinguished in distribution (Fig. 4.11). The concentration of Mg and Ca increased when the concentration of alkalis (Na+K) decreased. The groundwater samples were predominated by Ca-Cl, mixed Ca-Mg-Cl, mixed Ca-Mg-HCO_3, and Ca-HCO_3. The correlation matrix of physicochemical parameters in groundwater is shown in Fig. 4.12. Groundwater parameters were positively correlated with the normalized EWQI except for the pH. Out of these, the Na^+, HCO_3^-, SO_4^{2-}, TH, Ca^{2+}, EC, Mg^{2+}, Cl^-, and TDS had more than a correlation value of 0.5. The spatial distribution of

groundwater parameters is displayed in Fig. 4.13.

Table 4.11. Parameters relevant to groundwater samples

Parameters	Max	Min	Mean	SD	Standard	Entropy weight	P value	No. of samples out limit
EC (mS/cm)	6030.00	2.21	1409.32	995.51	2000 **	0.055	0.00	34
TH (mg/l)	1200.00	40.00	365.58	161.83	500 *	0.024	0.01	24
TDS (mg/l)	1720.00	75.00	587.86	261.04	1200 *	0.023	0.00	5
pH	9.00	5.60	7.27	0.40	7 - 8.5 *	0.001	0.00	38
Ca^{2+} (mg/l)	316.00	6.40	97.09	46.18	71 **	0.029	0.02	132
Mg^{2+} (mg/l)	150.00	0.20	29.96	21.02	30 **	0.055	0.00	70
Na^+ (mg/l)	340.00	0.00	64.82	66.36	200 **	0.110	0.00	10
K^+ (mg/l)	56.50	0.17	4.73	5.69	12 **	0.111	0.00	13
Cl^- (mg/l)	410.00	3.11	53.30	64.40	600 *	0.123	0.00	0
HCO_3^- (mg/l)	1200.00	23.35	461.94	172.55	120 **	0.019	0.01	173
SO_4^{2+} (mg/l)	585.11	0.26	42.39	82.73	250 *	0.219	0.00	5
NO_3^- (mg/l)	137.23	0.00	9.77	17.00	45 *	0.233	0.00	9

Note: * Standard value based on Thailand guidline, ** Standard value based on WHO.
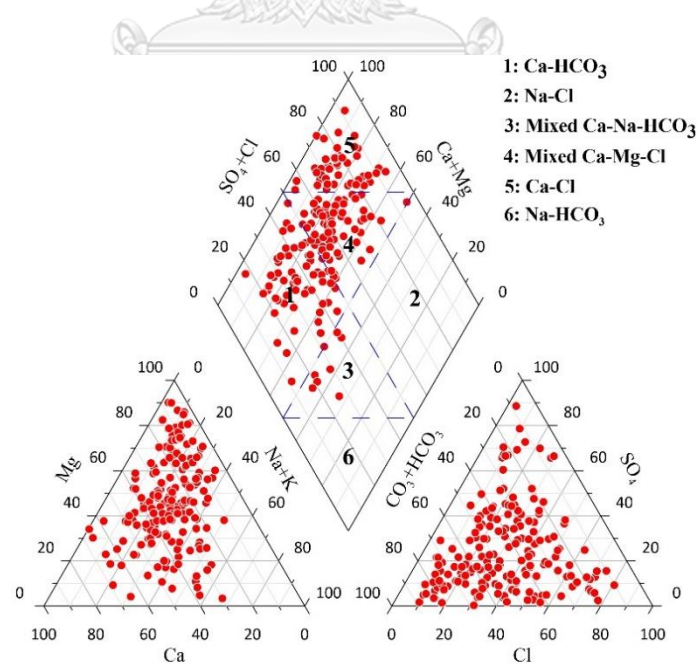
Fig. 4.11. Hydrogeochemical properties in groundwater samples
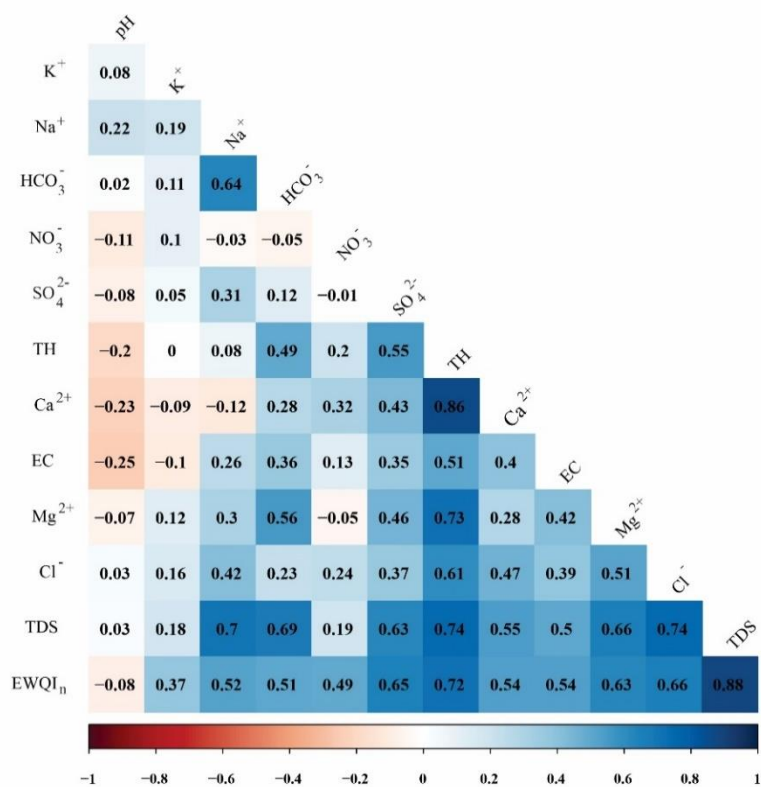
Fig. 4.12. Pearson correlation between groundwater parameters and the normalized EWQI
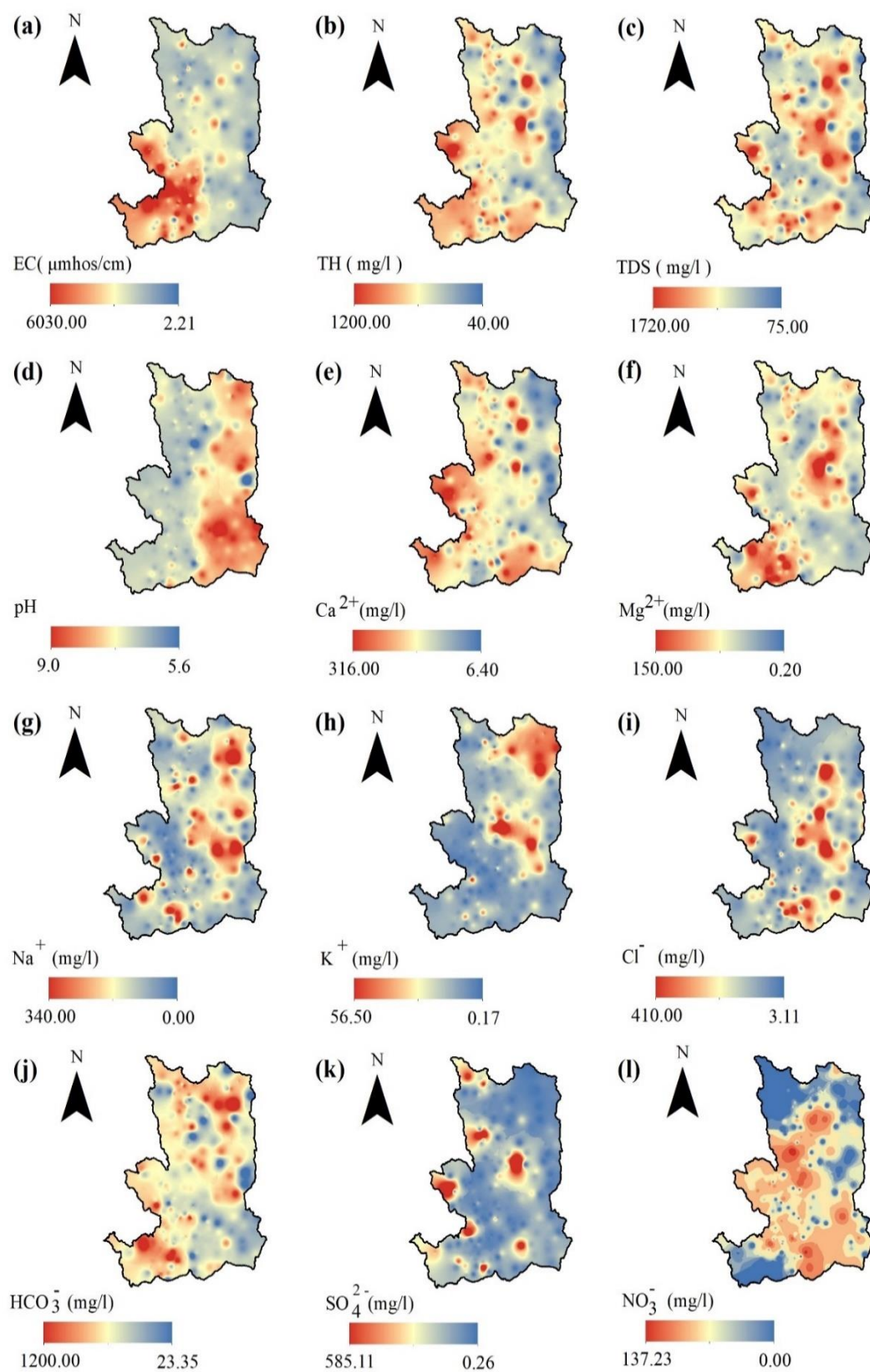
Fig. 4.13. Spatial distribution of groundwater parameters in the study area

4.3.2. Application of the RF and ANN with cross-validation and bootstrap in deciphering groundwater quality

The regression task of RF and ANN was conducted to predict the normalized EWQI in this study. Then, the normalized EWQI was deciphered into groundwater quality zones. The CV and B techniques were directly integrated into the RF and ANN algorithms, called RF-CV, RF-B, ANN-CV, and ANN-B models, respectively. For this task, the RF model was optimized from the error estimate to determine the n-tree (that is., the optimal number of trees). The optimal number of n-tree was 36 and 134 for the RF-CV and RF-B, respectively (Fig. 4.14). Compared to the RF, the ANN was established based on back-propagated feed-forward processes using the neural networks. Both the ANN-CV and ANN-B were optimal when the hidden layer was determined to be three nodes and a weight decay of 0.1 (Fig. 4.15). Fig. 4.16 presents the groundwater quality map using the RF-CV, RF-B, ANN-CV, and ANN-B models. For the RF-CV, 64.78% and 29.39% of the study area were good and very good groundwater quality, respectively. Meanwhile, only 0.58% and 0.08% were poor and very poor groundwater quality, respectively. The remaining 5.17% was designated to be moderate groundwater quality. For the RF-B, 62.92% and 30.24% of the study area were deciphered into good and very good groundwater quality. Moderate, poor, and very poor groundwater quality zones accounted for 6.23 %, 0.56%, and 0.05% of the whole region, respectively (Table 4.12). For the ANN-CV model, 41.94% and 52.66% of the total area were estimated to be of very good and good groundwater quality. Moderate and poor groundwater quality areas account for 4.50% and 0.90% of the total area, respectively. For the ANN-B, 53.46% and 39.90% of the total area were deciphered to be good and very good groundwater quality. Moderate, and poor zones constitute 5.67%, and 0.97% of the total area, respectively (Table 4.12).

Table 4.12. Statistical area in the models

Groundwater quality	RF-CV		RF-B		ANN-CV		ANN-B	
	Pixels	%	Pixels	%	Pixels	%	Pixels	%
Very good	2228061	29.39	2292578	30.24	3180026	41.94	3025163	39.90
Good	4911720	64.78	4771306	62.92	3992212	52.66	4053478	53.46
Moderate	391682	5.17	472007	6.23	341541	4.5	429246	5.67
Poor	43926	0.58	42205	0.56	67864	0.9	73756	0.97
Very poor	6254	0.08	3547	0.05	0	0.0	0	0.0
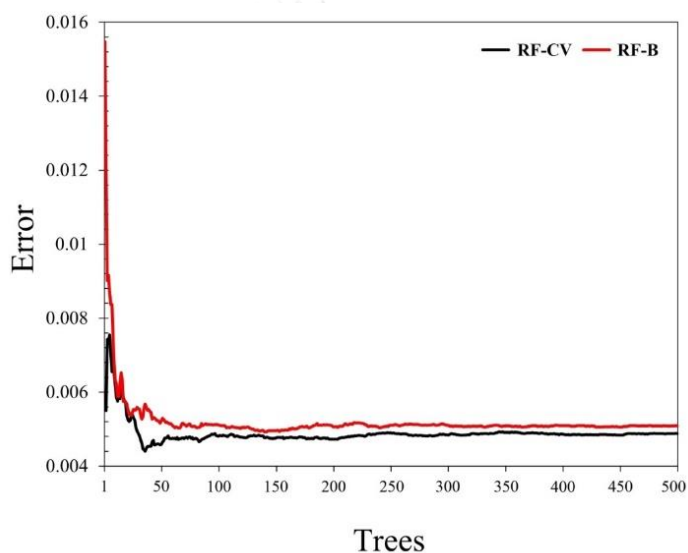
Fig. 4.14. Relationship between n-tree and predictive error in the RF-CV and RF-B models
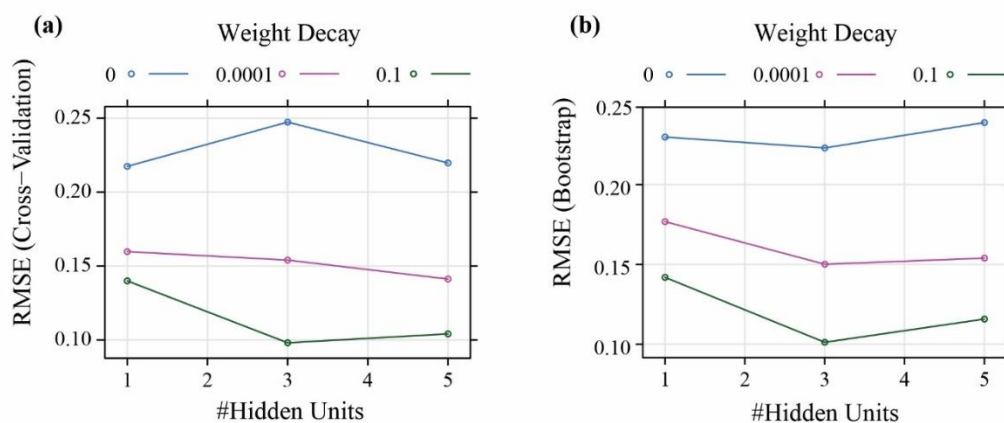
Fig. 4.15. Neural network of the ANN-CV (a) and ANN-B (b)
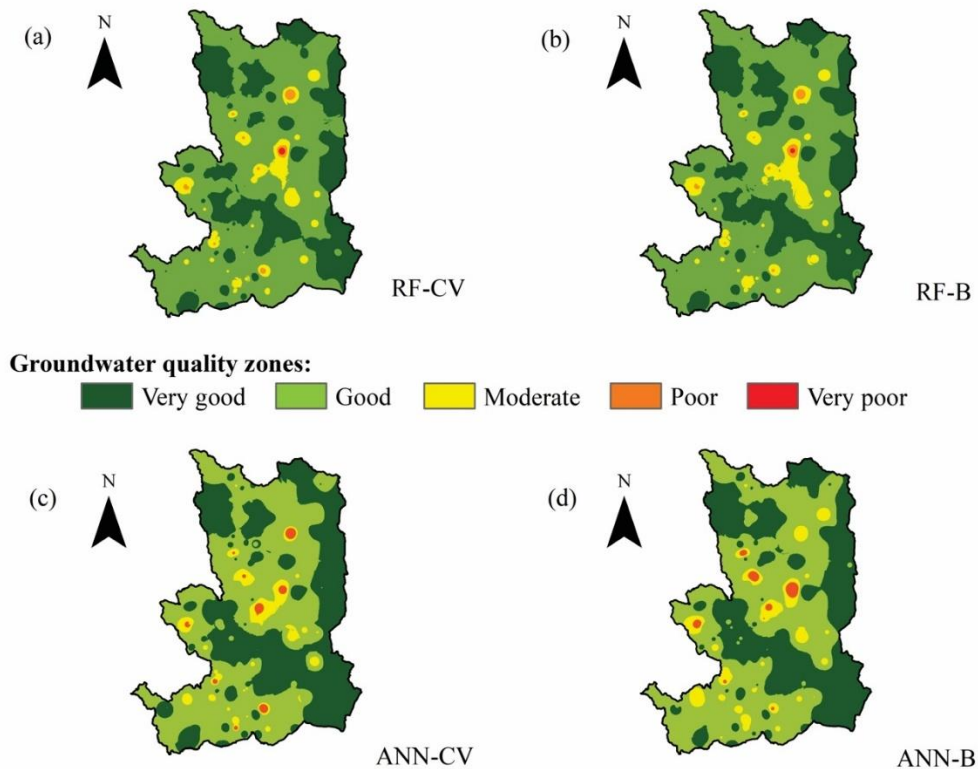
Fig. 4.16. Groundwater quality zones: RF-CV (a), RF-B (b), ANN-CV (c), ANN-B (d)

4.3.3. Validation

The validation process plays a key role to confirm the performance of machine learning models, which assists in choosing the best predicting model for mapping groundwater quality. Table 4.13 presents the validation indexes estimated by the models. Regarding the values of RMSE and MAE, the ANN-B model was the highest (RMSE = 0.101 and MAE = 0.065), followed by ANN-CV (RMSE = 0.098 and MAE = 0.061), RF-B (RMSE = 0.078 and MAE = 0.047), ANN-CV (RMSE = 0.067 and MAE = 0.043). However, the R^2 value of RF-CV was the highest (0.873) compared to the other three models. The RF-V's R^2 of 0.081 was second only to that of the RF-CV. The ANN-CV and ANN-B placed third and fourth when comparing the R^2 values with 0.709 and 0.617, respectively.

Table 4.13. Validation of the models

Models	RMSE	R²	MAE
RF-CV	0.067	0.873	0.043
RF-B	0.078	0.801	0.047
ANN-CV	0.098	0.709	0.061
ANN-B	0.101	0.617	0.065

4.3.4. Role of parameters in deciphering the groundwater quality

A stable and powerful model is normally decided by input variables (Şakar, 2018). Therefore, the importance of groundwater parameters needs revelation to decipher the groundwater quality maps, which helps to determine the contribution of physiochemical parameters to normalized EWQI. There was a great difference among models regarding the role of groundwater parameters in predicting the normalized EWQI (Table 4.14). In the RF models, TDS was the most important contribution to predicting the normalized EWQI (100) while the contribution of pH was zero. In the ANN-CV, the important parameters included pH (100), Na^+ (30.25%), Cl^- (29.54%), SO_4^{2-} (25.31%), Ca^{2+} (20.86%), and NO_3^- (20.79%). In the ANN-B, pH (100%), SO_4^{2-} (29.05%), and NO_3^- (28.39%) were the important parameters in deciphering groundwater quality.

Table 4.14. Contribution levels of groundwater parameters in deciphering the normalized EWQI

Model	RF-CV (%)	RF-B (%)	ANN-CV (%)	ANN-B (%)
Total dissolved solids	100.00	100.00	4.17	15.85
Chloride	0.31	15.62	29.54	8.94
Sulfate	4.85	14.15	25.31	29.05
Total hardness	1.17	13.25	12.58	4.46
Bicarbonate	2.54	10.95	1.87	13.13
Calcium	5.99	10.85	20.86	8.42
Nitrate	7.57	10.70	20.79	28.39
Sodium	0.75	5.61	30.25	7.53
Electric conductivity	2.32	4.36	0.00	0
Potassium	2.73	3.68	18.833	11.27
Magnesium	3.70	3.37	17.63	1.54
pH	0.00	0.00	100.00	100.00

4.3.5. Discussions

The protection of groundwater resources is a vital mission in many countries, especially where groundwater is used mainly for domestic and productive activities (Doveri et al., 2015). As a result, evaluating and delineating the groundwater quality zones is necessary to orient the groundwater use and management plan. For this reason, hydrologists have attempted to find an appropriate approach for delineating groundwater quality zones. This study pioneered integrating the CV and B techniques with two famous algorithms with high accuracy (RF and ANN). For this integration, four ML models were considered in deciphering groundwater quality in the eastern part of Kanchanaburi province, Thailand. The physicochemical parameter was referred to the previous publications in prestigious journals. In the comparison with previous studies (Adimalla, 2021; Amiri et al., 2014; Kumar and Augustine, 2022; Raheja et al., 2022; Singha et al., 2021), the normalized EWQI applied to delineate the groundwater quality zones, which contributes a general guideline for global groundwater investigations.

Our study indicated that there is a dramatic difference in the role of groundwater parameters in predicting the normalized EWQI in the four models, which can be explained by the nature of the approaches themselves. The RF models use decision trees to deal with data (Belgiu et al., 2016). While the ANN models are developed based on the neural network (Noori et al., 2010). In this analysis, our study found that TDS was the most important parameter in the RF models while pH was the most important parameter in the ANN models, which was not the same as a previous study by Singha et al. (2021). For this difference, the CV and B techniques were applied to the initial dataset while Singha et al. (2021) divided the initial dataset into two subsets with different ratios for the training and testing missions. This study indicated that the RF models were better than the ANN models in deciphering groundwater quality zones, which coincided with

some publications (Bui et al., 2020; Mallick et al., 2021; Nafouanti et al., 2021). In addition, the CV technique outperforms the B technique, which could be explained by the difference in building subsets (Brodeur et al., 2020). The CV technique uses the entire initial database while the B technique applies only a part of the initial database. Among four models, the most optimal model was the RF-CV (RMSE = 0.067, R^2 = 0.873, MAE = 0.043), followed by the RF-B (RMSE = 0.078, R^2 = 0.801, MAE = 0.047), ANN-CV (RMSE = 0.098, R^2 = 0.709, MAE = 0.061), and ANN-B (RMSE = 0.101, R^2 = 0.671, MAE = 0.065). As a result, this study proposes the RF-CV model could apply to map groundwater quality not only in Thailand but also in other parts of the world. Our study also indicated that the performance of the RF models depends on the number of trees (Bernard et al., 2009). While the performance of the RF models depends on the number of intermediate nodes (Zannou et al., 2021).

According to the RF-CV model, approximately 94% of the study's area was good to very good groundwater quality. The rest of the territory (about 6%) was medium, poor, and very poor groundwater quality, which was distributed in the middle regions of the study area. In the comparison with the land use map, poor, and very poor groundwater quality occurred in the agricultural and urban regions, which agrees with a study by Sajedi-Hosseini et al. (2018). Agricultural activities (like lime and inorganic fertilizer practices), limestone quarries, and geological settings were the main reasons leading to poor quality in the study area (Deshmukh, 2013; Ghaffari et al., 2021; Saha et al., 2019; Sridharan and Senthil Nathan, 2017). This study suggests that zones with good and very good groundwater quality can be used for daily purposes of local people while the zones classified very poor, poor and moderate groundwater quality should be used for agricultural or industrial purposes. However, the calcium, magnesium, and bicarbonate in groundwater should be handled before drinking because our physicochemical analysis showed that groundwater in the study area was mainly polluted by $Ca^{2+}, HCO_3^-, Mg^{2+}$.

Chapter 5. CONCLUSION AND SUSGESTION

5.1. Conclusion

The AHP, FR, and RF models were used in this study to represent the expert decision, statistical, and machine learning approaches for the GWPZ mapping method comparison. In particular, the study aimed to map the potential of achieving a groundwater yield above 10 m3/h for Kanchanaburi Province in Thailand. An ensemble of the three models was also created and compared to the individual models in the mapping of GWPZ. The results showed that all models achieved similarly good prediction with an AUC value of 0.72, 0.74, 0.76, and 0.80 for the AHP, FR, RF, and ensemble models, respectively. The ensemble model improved the overall predictive power and performed better in areas where all the individual models failed. This study illustrated the usefulness of the ensemble of models from different approaches in mapping GWPZ and discussed how future studies can adopt similar approaches to improve the delineation of GWPZ for policy foci in sustainable groundwater management. This study recommends that the GWPZ map achieved by the ensemble model be integrated with the surface water planning map, land use planning map, and economic map to assist scientists, policymakers, and managers in Kanchanaburi Province, Thailand, in inclusive and sustainable development.

This study attempted to delineate the Ni contamination risk zones using a hybrid model of the Maxent and AHP modeling approach. This model works well with an AUC value of 0.86 and an accuracy of 0.85 in delineating the geographical distribution of Ni contamination risk in groundwater. Eight factors, including altitude, distance to roads, distance to waterbodies, geology, land use, rainfall, slope, and soil type, were selected to set up the model. Among the eight investigated factors in this study, land use was the most relevant factor for identifying possibly contaminated locations, followed by soil type. The urban area was the most likely Ni

contamination in groundwater. The map of Ni contamination risk zones in groundwater could provide information for residents and decision-makers in using groundwater resources for various purposes. It could help Thailand's groundwater resource management agencies grasp the practical basis and find solutions for the protection of groundwater resources in areas with high Ni contamination risk.

This study evaluated groundwater quality in the eastern part of Kanchanaburi province and developed four ML models by integrated the RF and ANN algorithms with the CV and B techniques. The EWQI was converted into the normalized EWQI. The study found that groundwater quality in the study area was polluted with calcium, magnesium, and bicarbonate and that the RF-CV was the best in deciphering the groundwater quality map, compared to the RF-B, ANN-CV, and ANN-B. In the RF-CV, TDS was the most important parameter while other parameters were little contribution to predicting the normalized EWQI. The groundwater quality zones were classified into five levels in the study area. Regarding the best model (RF-CV), poor and very poor groundwater quality occurred in the agricultural areas. The results from the present study can be a reference document for developing and mapping groundwater quality in the future.

5.2. Suggestion

For mapping groundwater yield potential, the AHP, FR, and RF models were used in this case study to represent the expert decision, statistical, and machine learning approaches in GWPZ mapping. However, this study acknowledges that these approaches are broad topics on their own and thus this study's comparison of the particular models might not be representative. This study strongly encourages future studies to explore different model combinations to fortify this study's initial findings, as this study pioneered the research. For studies that wish to follow this study's approach to GWPZ mapping, this study recommends that they first employ the AHP

methods to generate the GWPZ map and then collect a small set of ground truth data to validate the model results. If the results do not meet expectations, more ground truth data can be collected to create the FR model, and if further accuracy improvement is required, much more data should be collected to build the RF model. As illustrated by this study's analysis, one should always attempt ensemble models when data is available to improve the overall predictive power. Additionally, this study also suggests that groundwater resources in the Western part of Kanchanaburi, including Sangkhla Buri, Thong Pha Phum, Si Sawat, Sai Yok districts, should be strictly protected due to the low groundwater yield potential and that in the Eastern part of the province, including Nong Prue, Lao Khan, Bo Phloi, Huai Krachrao, Muong Kanchanaburi, Tha Muang, Tha Maka, Dan Makham districts, can exploit for purposes of socio-economic development.

For mapping contamination risk in groundwater, this study recommends that the hybrid model between Maxent and AHP can also be applied to investigate other heavy metal elements in groundwater because of its own advantages. The input influencing factors of the model could be easily collected from the field, government agencies, and free satellite images. Consequently, budget, human labor, and time could be saved using this process. However, Ni concentration depends on the accuracy of the analysis instrument. Although the modeling approach for mapping the Ni contamination risk zone gives a good result, groundwater is constantly moving, so the groundwater quality at different times will be different. Therefore, the assessment of Ni content in groundwater should be carried out regularly in the future. Additionally, the present model still has limitations in terms of the data on solid wastes, wastewater, waste gas, surface water pollution, and soil contamination. Hence, these influencing factors should be used to control the future model. This study also recommends local people in areas with a high risk of groundwater pollution, especially urban regions, should not use groundwater for

daily drinking purposes.

For mapping groundwater quality, this study suggests that the RF-CV model should be applied to map groundwater quality not only in Thailand but also in other parts of the world. However, this study focuses only on the physicochemical parameters to estimate the EWQI, the study recommends that future studies should consider the heavy metal elements in deciphering groundwater quality maps. Additionally, this study limits in considering the RF and ANN combined with the cross-validation and bootstrap, therefore an exploration of the performance of other machine learning is needed to be conducted in the future to determine the best model for deciphering groundwater quality map. Simultaneously, groundwater samples were collected at a time due to the limitation of financial support, therefore, the seasonal variations of groundwater quality are not considered, which can affect the results.

From the results, groundwater agencies can release policies on groundwater. It can be done by publicizing the list of restricted areas from the exploitation of groundwater and orienting in reasonable groundwater exploitation and usage with different purposes.

REFERENCES

Abd Manap, M., Nampak, H., Pradhan, B., Lee, S., Sulaiman, W.N.A., Ramli, M. F. (2014). Application of probabilistic-based frequency ratio model in groundwater potential mapping using remote sensing data and GIS. *Arabian Journal of Geosciences*, *7*(2), 711-724. 10.1007/s12517-012-0795-z.

Abd Manap, M., Sulaiman, W.N.A., Ramli, M.F., Pradhan, B., Surip, N. (2013). A knowledge-driven GIS modeling technique for groundwater potential mapping at the Upper Langat Basin, Malaysia. *Arabian Journal of Geosciences*, *6*(5), 1621-1637. 10.1007/s12517-011-0469-2.

Abdalla, F. (2012). Mapping of groundwater prospective zones using remote sensing and GIS techniques: A case study from the Central Eastern Desert, Egypt. *Journal of African Earth Sciences*, *70*, 8-17. https://doi.org/10.1016/j.jafrearsci.2012.05.003.

Abijith, D., Saravanan, S., Singh, L., Jennifer, J. J., Saranya, T., Parthasarathy, K. S. S. (2020). GIS-based multi-criteria analysis for identification of potential groundwater recharge zones-a case study from Ponnaniyaru watershed, Tamil Nadu, India. *HydroResearch*, *3*, 1-14. https://doi.org/10.1016/j.hydres.2020.02.002.

Achu, A. L., Thomas, J., Reghunath, R. (2020). Multi-criteria decision analysis for delineation of groundwater potential zones in a tropical river basin using remote sensing, GIS and analytical hierarchy process (AHP). *Groundwater for Sustainable Development*, *10*, 100365. https://doi.org/10.1016/j.gsd.2020.100365.

Adiat, K. A. N., Nawawi, M. N. M., Abdullah, K. (2012). Assessing the accuracy of GIS-based elementary multi criteria decision analysis as a spatial prediction tool–a case of predicting potential zones of sustainable groundwater resources. *Journal of Hydrology, 440,* 75-89.

https://doi.org/10.1016/j.jhydrol.2012.03.028.

Adimalla, N. (2021). Application of the entropy weighted water quality index (EWQI) and the pollution index of groundwater (PIG) to assess groundwater quality for drinking purposes: a case study in a rural area of Telangana State, India. *Archives of environmental contamination and toxicology*, *80*(1), 31-40. https://doi.org/10.1007/s00244-020-00800-4.

Agarwal, E., Agarwal, R., Garg, R. D., & Garg, P. K. (2013). Delineation of groundwater potential zone: an AHP/ANP approach. *Journal of earth system science*, *122*(3), 887-898.

Agarwal, R., & Garg, P. K. (2016). Remote sensing and GIS based groundwater potential & recharge zones mapping using multi-criteria decision making technique. *Water resources management*, *30*(1), 243-260. 10.1007/s11269-015-1159-8.

Agrawal, P., Sinha, A., Kumar, S., Agarwal, A., Banerjee, A., Villuri, V. G. K., Pasupuleti, S., Kumar, D., Kaushal, D. R., Gosain, K. A., & Pasupuleti, S. (2021). Exploring artificial intelligence techniques for groundwater quality assessment. *Water*, *13*(9), 1172. https://doi.org/10.3390/w13091172.

Ahmad, M. W., Mourshed, M., & Rezgui, Y. (2017). Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and buildings*, *147*, 77-89. https://doi.org/10.1016/j.enbuild.2017.04.038.

Ahmed, K., Shahid, S., Bin Harun, S., Ismail, T., Nawaz, N., & Shamsudin, S. (2015). Assessment of groundwater potential zones in an arid region based on catastrophe theory. *Earth Science Informatics*, *8*(3), 539-549. 10.1007/s12145-014-0173-3.

Ahmed, R., & Sajjad, H. (2018). Analyzing factors of groundwater potential and its

relation with population in the Lower Barpani Watershed, Assam, India. *Natural Resources Research*, *27*(4), 503-515. https://doi.org/10.1007/s11053-017-9367-y.

Aizebeokhai, A. P., Olayinka, A. I., & Singh, V. S. (2010). Application of 2D and 3D geoelectrical resistivity imaging for engineering site investigation in a crystalline basement terrain, southwestern Nigeria. *Environmental Earth Sciences*, *61*(7), 1481-1492. 10.1007/s12665-010-0464-z.

Akinluyi, F. O., Olorunfemi, M. O., & Bayowa, O. G. (2018). Investigation of the influence of lineaments, lineament intersections and geology on groundwater yield in the basement complex terrain of Ondo State, Southwestern Nigeria. *Applied water science*, *8*(1), 1-13. https://doi.org/10.1007/s13201-018-0686-x.

Al-Abadi, A. M. (2017a). Modeling of groundwater productivity in northeastern Wasit Governorate, Iraq using frequency ratio and Shannon's entropy Models. *Applied Water Science*, *7*(2), 699-716. 10.1007/s13201-015-0283-1.

Al-Abadi, A. M. (2015). Groundwater potential mapping at northeastern Wasit and Missan governorates, Iraq using a data-driven weights of evidence technique in framework of GIS. *Environmental earth sciences*, *74*(2), 1109-1124. 10.1007/s12665-015-4097-0.

Al-Abadi, A. M. (2017b). A novel geographical information system-based Ant Miner algorithm model for delineating groundwater flowing artesian well boundary: a case study from Iraqi southern and western deserts. *Environmental Earth Sciences*, *76*(15), 1-19. 10.1007/s12665-017-6876-2.

Al-Djazouli, M. O., Elmorabiti, K., Rahimi, A., Amellah, O., & Fadil, O. A. M. (2021). Delineating of groundwater potential zones based on remote sensing, GIS and analytical hierarchical process: a case of Waddai, eastern Chad. *GeoJournal*, *86*(4), 1881-1894. https://doi.org/10.1007/s10708-020-10160-

0.

Al-Ruzouq, R., Shanableh, A., Merabtene, T., Siddique, M., Khalil, M. A., Idris, A., & Almulla, E. (2019). Potential groundwater zone mapping based on geo-hydrological considerations and multi-criteria spatial analysis: North UAE. *Catena*, *173*, 511-524. https://doi.org/10.1016/j.catena.2018.10.037.

Ali, A., Farid, H. U., & Khan, M. M. H. (2020). Divergent effect of rainfall, temperature and surface water bodies on groundwater quality in Haveli Canal Circle of Multan Irrigation Zone, Southern Punjab. *Pakistan. Journal of Environmental and Agricultural Sciences*, *22*(4), 25-36.

Ali, H. Y., Priju, C. P., & Prasad, N. N. (2015). Delineation of groundwater potential zones in deep midland aquifers along Bharathapuzha river basin, Kerala using geophysical methods. *Aquatic Procedia*, *4*, 1039-1046. https://doi.org/10.1016/j.aqpro.2015.02.131.

Alizamir, M., Kisi, O., & Zounemat-Kermani, M. (2018). Modelling long-term groundwater fluctuations by extreme learning machine using hydro-climatic data. *Hydrological sciences journal*, *63*(1), 63-73. https://doi.org/10.1080/02626667.2017.1410891.

Alizamir, M., Sobhanardakani, S., & Shahrabadi, A. H. (2019). Prediction of heavy metals concentration in the groundwater resources in Razan Plain: extreme learning machine vs. artificial neural network and multivariate adaptive regression spline. *Annals of Military and Health Sciences Research*, *17*(4). 10.5812/amh.98554.

Alonso, J. A., & Lamata, M. T. (2006). Consistency in the analytic hierarchy process: a new approach. *International journal of uncertainty, fuzziness and knowledge-based systems*, *14*(04), 445-459. https://doi.org/10.1142/S0218488506004114.

Amiri, V., Rezaei, M., & Sohrabi, N. (2014). Groundwater quality assessment using entropy weighted water quality index (EWQI) in Lenjanat, Iran. *Environmental Earth Sciences*, *72*(9), 3479-3490. 10.1007/s12665-014-3255-0.

Andermann, C., Longuevergne, L., Bonnet, S., Crave, A., Davy, P., & Gloaguen, R. (2012). Impact of transient groundwater storage on the discharge of Himalayan rivers. *Nature geoscience*, *5*(2), 127-132. 10.1038/NGEO1356.

Anderson, E. I., & Bakker, M. (2008). Groundwater flow through anisotropic fault zones in multiaquifer systems. *Water Resources Research*, *44*(11). https://doi.org/10.1029/2008WR006925.

Andualem, T. G., & Demeke, G. G. (2019). Groundwater potential assessment using GIS and remote sensing: A case study of Guna tana landscape, upper blue Nile Basin, Ethiopia. *Journal of Hydrology: Regional Studies*, *24*, 100610. https://doi.org/10.1016/j.ejrh.2019.100610.

Arabameri, A., Rezaei, K., Cerda, A., Lombardo, L., & Rodrigo-Comino, J. (2019). GIS-based groundwater potential mapping in Shahroud plain, Iran. A comparison among statistical (bivariate and multivariate), data mining and MCDM approaches. *Science of the total environment*, *658*, 160-177. https://doi.org/10.1016/j.scitotenv.2018.12.115.

Arkoprovo, B., Adarsa, J., & Prakash, S. S. (2012). Delineation of groundwater potential zones using satellite remote sensing and geographic information system techniques: a case study from Ganjam district, Orissa, India.

Arulbalaji, P., Padmalal, D., & Sreelash, K. (2019). GIS and AHP techniques based delineation of groundwater potential zones: a case study from southern Western Ghats, India. *Scientific reports*, *9*(1), 1-17. https://doi.org/10.1038/s41598-019-38567-x.

Arunbose, S., Srinivas, Y., Rajkumar, S., Nair, N. C., & Kaliraj, S. (2021). Remote sensing, GIS and AHP techniques based investigation of groundwater potential zones in the Karumeniyar river basin, Tamil Nadu, southern India. *Groundwater for Sustainable Development*, *14*, 100586. https://doi.org/10.1016/j.gsd.2021.100586.

Asadi, E., Isazadeh, M., Samadianfard, S., Ramli, M. F., Mosavi, A., Nabipour, N., ... & Chau, K. W. (2019). Groundwater quality assessment for sustainable drinking and irrigation. *Sustainability*, *12*(1), 177. https://doi.org/10.3390/su12010177.

Assatse, W. T., Nouck, P. N., Tabod, C. T., Akame, J. M., & Biringanine, G. N. (2016). Hydrogeological activity of lineaments in Yaoundé Cameroon region using remote sensing and GIS techniques. *The Egyptian Journal of Remote Sensing and Space Science*, *19*(1), 49-60. https://doi.org/10.1016/j.ejrs.2015.12.006.

Avinash, K., Jayappa, K. S., & Deepika, B. (2011). Prioritization of sub-basins based on geomorphology and morphometricanalysis using remote sensing and geographic informationsystem (GIS) techniques. *Geocarto International*, *26*(7), 569-592. https://doi.org/10.1080/10106049.2011.606925.

Avtar, R., Singh, C. K., Shashtri, S., Singh, A., & Mukherjee, S. (2010). Identification and analysis of groundwater potential zones in Ken–Betwa river linking area using remote sensing and geographic information system. *Geocarto International*, *25*(5), 379-396. https://doi.org/10.1080/10106041003731318.

Azimi, S., Azhdary Moghaddam, M., & Hashemi Monfared, S. A. (2018). Spatial assessment of the potential of groundwater quality using fuzzy AHP in GIS. *Arabian Journal of Geosciences*, *11*(7), 1-22. https://doi.org/10.1007/s12517-018-3484-8.

Bagyaraj, M., Mengistie, A. T., Gnanachandrasamy, G., & Gemechu, B. (2019). Data of remote sensing and GIS-to demarcate the potential sector of groundwater in

Debre Berhan, Amhara region, Ethiopia. *Data in brief*, *26*. https://doi.org/10.1016/j.dib.2019.104542.

Bagyaraj, M., Ramkumar, T., Venkatramanan, S., & Gurugnanam, B. (2013). Application of remote sensing and GIS analysis for identifying groundwater potential zone in parts of Kodaikanal Taluk, South India. *Frontiers of Earth Science*, *7*(1), 65-75. 10.1007/s11707-012-0347-6.

Barzegar, R., Moghaddam, A. A., Deo, R., Fijani, E., & Tziritis, E. (2018). Mapping groundwater contamination risk of multiple aquifers using multi-model ensemble of machine learning algorithms. *Science of the total environment*, *621*, 697-712. https://doi.org/10.1016/j.scitotenv.2017.11.185.

Basavaraj, H., & Nijagunappa, R. (2011). Identification of groundwater potential zone using geoinformatics in Ghataprabha basin, North Karnataka, India. *international Journal of Geomatics and Geosciences*, *2*(1), 91-109.

Batte, A. G., Muwanga, A., & Sigrist, W. P. (2008). Evaluating the use of vertical electrical sounding as a groundwater exploration technique to improve on the certainty of borehole yield in Kamuli District (Eastern Uganda). *AJST*, *9*(1), 73.

Baumann, T., Fruhstorfer, P., Klein, T., & Niessner, R. (2006). Colloid and heavy metal transport at landfill sites in direct contact with groundwater. *Water research*, *40*(14), 2776-2786. https://doi.org/10.1016/j.watres.2006.04.049.

Beck, R. A. (2003). Remote sensing and GIS as counterterrorism tools in the Afghanistan war: A case study of the Zhawar Kili region. *The Professional Geographer*, *55*(2), 170-179.

Bedi, S., Samal, A., Ray, C., & Snow, D. (2020). Comparative evaluation of machine learning models for groundwater quality assessment. *Environmental Monitoring and Assessment*, *192*(12), 1-23. https://doi.org/10.1007/s10661-020-08695-3.

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, *114*, 24-31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

Bense, V. F., & Person, M. A. (2006). Faults as conduit-barrier systems to fluid flow in siliciclastic sedimentary aquifers. *Water Resources Research*, *42*(5). https://doi.org/10.1029/2005WR004480.

Bernard, S., Heutte, L., & Adam, S. (2009). On the selection of decision trees in random forests. In *2009 International Joint Conference on Neural Networks* (pp. 302-307). IEEE. 10.1109/IJCNN.2009.5178693.

Berrar, D. (2019). Cross-Validation.

Boateng, T. K., Opoku, F., Acquaah, S. O., & Akoto, O. (2016). Groundwater quality assessment using statistical approach and water quality index in Ejisu-Juaben Municipality, Ghana. *Environmental Earth Sciences*, *75*(6), 1-14. 10.1007/s12665-015-5105-0.

Bonham-Carter, G. F., (1994). *Geographic information systems for geoscientists: modelling with GIS* (No. 13). Elsevier.

Boonkaewwan, S., Sonthiphand, P., & Chotpantarat, S. (2020). Mechanisms of arsenic contamination associated with hydrochemical characteristics in coastal alluvial aquifers using multivariate statistical technique and hydrogeochemical modeling: A case study in Rayong province, eastern Thailand. *Environmental Geochemistry and Health*, *43*(1), 537-566. https://doi.org/10.1007/s10653-020-00728-7.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Brodeur, Z. P., Herman, J. D., & Steinschneider, S. (2020). Bootstrap aggregation and cross-validation methods to reduce overfitting in reservoir control policy

search. *Water Resources Research*, *56*(8), e2020WR027184. https://doi.org/10.1029/2020WR027184.

Bui, D. T., Khosravi, K., Karimi, M., Busico, G., Khozani, Z. S., Nguyen, H., Mastrocicco, M., Tedesco, D., Cuoco, E., & Kazakis, N. (2020). Enhancing nitrate and strontium concentration prediction in groundwater by using new data mining algorithm. *Science of the Total Environment*, *715*, 136836. https://doi.org/10.1016/j.scitotenv.2020.136836.

Bunopas, S. (1976). Geologic Map of Changwat Suphan Buri, Sheet ND 47-7, Scale 1: 250,000. *Report of Investigation*, (16).

Catani, F., Lagomarsino, D., Segoni, S., & Tofani, V. (2013). Exploring model sensitivity issues across different scales in landslide susceptibility. *Nat Hazards Earth Syst Sci*, *13*, 2815-2831. 10.5194/nhessd-1-583-2013.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific model development*, *7*(3), 1247-1250. https://doi.org/10.5194/gmd-7-1247-2014.

Chakraborty, S., & Kumar, R. N. (2016). Assessment of groundwater quality at a MSW landfill site using standard and AHP based water quality index: a case study from Ranchi, Jharkhand, India. *Environmental monitoring and assessment*, *188*(6), 1-18. 10.1007/s10661-016-5336-x.

Chaminé, H. I., Carvalho, J. M., Teixeira, J., & Freitas, L. (2015). Role of hydrogeological mapping in groundwater practice: back to basics. *European Geologist Journal*, *40*, 34-42. http://hdl.handle.net/10400.22/18051.

Chaudhary, B. S., & Kumar, S. (2018). Identification of groundwater potential zones using remote sensing and GIS of KJ Watershed, India. *Journal of the Geological Society of India*, *91*(6), 717-721. 10.1007/s12594-018-0929-3.

Chen, W., Li, H., Hou, E., Wang, S., Wang, G., Panahi, M., Li, T., Penf, T., Guo, C., Niu, C., Xiao, L., Wang, J., & Ahmad, B. B. (2018). GIS-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models. *Science of the Total Environment*, *634*, 853-867. https://doi.org/10.1016/j.scitotenv.2018.04.055.

Chen, W., Panahi, M., Khosravi, K., Pourghasemi, H. R., Rezaie, F., & Parvinnezhad, D. (2019a). Spatial prediction of groundwater potentiality using ANFIS ensembled with teaching-learning-based and biogeography-based optimization. *Journal of Hydrology*, *572*, 435-448. https://doi.org/10.1016/j.jhydrol.2019.03.013.

Chen, W., Pradhan, B., Li, S., Shahabi, H., Rizeei, H. M., Hou, E., & Wang, S. (2019b). Novel hybrid integration approach of bagging-based fisher's linear discriminant function for groundwater potential analysis. *Natural Resources Research*, *28*(4), 1239-1258. https://doi.org/10.1007/s11053-019-09465-w.

Chen, W., Tsangaratos, P., Ilia, I., Duan, Z., & Chen, X. (2019c). Groundwater spring potential mapping using population-based evolutionary algorithms and data mining methods. *Science of The Total Environment*, *684*, 31-49. https://doi.org/10.1016/j.scitotenv.2019.05.312.

Chen, Z., Wu, G., Wu, Y., Wu, Q., Shi, Q., Ngo, H. H., Saucedo, O. A. V., & Hu, H. Y. (2020). Water Eco-Nexus Cycle System (WaterEcoNet) as a key solution for water shortage and water environment problems in urban areas. *Water Cycle*, *1*, 71-77. https://doi.org/10.1016/j.watcyc.2020.05.004.

Cheng, Y. S., Yu, T. T., & Son, N. T. (2021). Random Forests for Landslide Prediction in Tsengwen River Watershed, Central Taiwan. *Remote Sensing*, *13*(2), 199. https://doi.org/10.3390/rs13020199.

Chernick, M. R. (2012). Resampling methods. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *2*(3), 255-262.

https://doi.org/10.1002/widm.1054.

Choi, J., Oh, H. J., Lee, H. J., Lee, C., & Lee, S. (2012). Combining landslide susceptibility maps obtained from frequency ratio, logistic regression, and artificial neural network models using ASTER images and GIS. *Engineering geology*, *124*, 12-23. https://doi.org/10.1016/j.enggeo.2011.09.011.

Chotpantarat, S., & Boonkaewwan, S. (2018). Impacts of land-use changes on watershed discharge and water quality in a large intensive agricultural area in Thailand. *Hydrological sciences journal*, *63*(9), 1386-1407. https://doi.org/10.1080/02626667.2018.1506128.

Chotpantarat, S., Parkchai, T., & Wisitthammasri, W. (2020). Multivariate statistical analysis of hydrochemical data and stable isotopes of groundwater contaminated with nitrate at Huay Sai Royal Development Study Center and adjacent areas in Phetchaburi Province, Thailand. *Water*, *12*(4), 1127. https://doi.org/10.3390/w12041127.

Chotpantarat, S., & Thamrongsrisakul, J. (2021). Natural and anthropogenic factors influencing hydrochemical characteristics and heavy metals in groundwater surrounding a gold mine, Thailand. *Journal of Asian Earth Sciences*, *211*, 104692. https://doi.org/10.1016/j.jseaes.2021.104692.

Chotpantarat, S., Wongsasuluk, P., Siriwong, W., Borjan, M., & Robson, M. (2014). Non-carcinogenic hazard maps of heavy metal contamination in shallow groundwater for adult and aging populations at an agricultural area in northeastern Thailand. *Human and Ecological Risk Assessment: An International Journal*, *20*(3), 689-703. https://doi.org/10.1080/10807039.2013.832998.

Chowdhury, A., Jha, M. K., Chowdary, V. M., & Mal, B. C. (2009). Integrated remote sensing and GIS-based approach for assessing groundwater potential in West

Medinipur district, West Bengal, India. *International Journal of Remote Sensing*, *30*(1), 231-250. https://doi.org/10.1080/01431160802270131.

Chuaviroj, S.J.B.G.S.D., Department of Mineral Resources, 1991. *Geotectonic of Thailand*. 58

Chuma, C., Hlatywayo, D. J., Orimoogunje, O. O., & Akinyede, J. O. (2013). Application of remote sensing and geographical information systems in determining the groundwater potential in the crystalline basement of Bulawayo metropolitan area, Zimbabwe. 10.4236/ars.2013.22019.

Condon, L. E., & Maxwell, R. M. (2015). Evaluating the relationship between topography and groundwater using outputs from a continental‐scale integrated hydrology model. *Water Resources Research*, *51*(8), 6602-6621. https://doi.org/10.1002/2014WR016774.

Cortadellas, T., Argacha, P., Acosta, J., Rabasa, J., Peiró, R., Gomez, M., Navarro-Golobart, A., Sanchez-Mendez, S., Martinez-Medina, M., Botey, M., Munoz-Ramos, C., & Xiberta, M. (2017). Estimation of tumor size in breast cancer comparing clinical examination, mammography, ultrasound and MRI—correlation with the pathological analysis of the surgical specimen. *Gland surgery*, *6*(4), 330.

Da Lio, C., Tosi, L., Zambon, G., Vianello, A., Baldin, G., Lorenzetti, G., Manfe, G., & Teatini, P. (2013). Long-term groundwater dynamics in the coastal confined aquifers of Venice (Italy). *Estuarine, Coastal and Shelf Science*, *135*, 248-259. https://doi.org/10.1016/j.ecss.2013.10.021.

Dai, A. (2013). Increasing drought under global warming in observations and models. *Nature climate change*, *3*(1), 52-58. 10.1038/NCLIMATE1633.

Dar, I. A., Sankar, K., & Dar, M. A. (2010). Remote sensing technology and geographic

information system modeling: an integrated approach towards the mapping of groundwater potential zones in Hardrock terrain, Mamundiyar basin. *Journal of Hydrology*, *394*(3-4), 285-295. https://doi.org/10.1016/j.jhydrol.2010.08.022.

Dar, I. A., Sankar, K., & Dar, M. A. (2011). Deciphering groundwater potential zones in hard rock terrain using geospatial technology. *Environmental monitoring and assessment*, *173*(1), 597-610. 10.1007/s10661-010-1407-6.

Dar, T., Rai, N., & Bhat, A. (2021). Delineation of potential groundwater recharge zones using analytical hierarchy process (AHP). *Geology, Ecology, and Landscapes*, *5*(4), 292-307. https://doi.org/10.1080/24749508.2020.1726562.

Das, B., Pal, S. C., Malik, S., & Chakrabortty, R. (2019). Modeling groundwater potential zones of Puruliya district, West Bengal, India using remote sensing and GIS techniques. *Geology, Ecology, and Landscapes*, *3*(3), 223-237. https://doi.org/10.1080/24749508.2018.1555740.

Das, S., Pardeshi, S. D., Kulkarni, P. P., & Doke, A. (2018). Extraction of lineaments from different azimuth angles using geospatial techniques: a case study of Pravara basin, Maharashtra, India. *Arabian Journal of Geosciences*, *11*(8), 1-13. https://doi.org/10.1007/s12517-018-3522-6.

Das, S. (2019). Comparison among influencing factor, frequency ratio, and analytical hierarchy process techniques for groundwater potential zonation in Vaitarna basin, Maharashtra, India. *Groundwater for Sustainable Development*, *8*, 617-629. https://doi.org/10.1016/j.gsd.2019.03.003.

Deepika, B., Avinash, K., & Jayappa, K. S. (2013). Integration of hydrological factors and demarcation of groundwater prospect zones: insights from remote sensing and GIS techniques. *Environmental earth sciences*, *70*(3), 1319-1338. 10.1007/s12665-013-2218-1.

Deshmukh, K. K. (2013). Impact of human activities on the quality of groundwater from Sangamner Area, Ahmednagar District, Maharashtra, India. *International Research Journal of Environment Sciences*, *2*(8), 66-74.

DGR, 2006. Nationwide Groundwater Exploration and Mapping of Limestone and Hard Rock Aquifers Project.

Dhar, A., Sahoo, S., & Sahoo, M. (2015). Identification of groundwater potential zones considering water quality aspect. *Environmental Earth Sciences*, *74*(7), 5663-5675. 10.1007/s12665-015-4580-7.

Díaz-Alcaide, S., & Martínez-Santos, P. (2019). Advances in groundwater potential mapping. *Hydrogeology Journal*, *27*(7), 2307-2324. https://doi.org/10.1007/s10040-019-02001-3.

Doveri, M., Menichini, M., & Scozzari, A. (2015). Protection of groundwater resources: worldwide regulations and scientific approaches. *Threats to the Quality of Groundwater Resources*, 13-30. https://doi.org/10.1007/698_2015_421.

Edet, A. E., Okereke, C. S., Teme, S. C., & Esu, E. O. (1998). Application of remote-sensing data to groundwater exploration: a case study of the Cross River State, southeastern Nigeria. *Hydrogeology journal*, *6*(3), 394-404.

Egbueri, J. C. (2020). Heavy metals pollution source identification and probabilistic health risk assessment of shallow groundwater in Onitsha, Nigeria. *Analytical Letters*, *53*(10), 1620-1638. https://doi.org/10.1080/00032719.2020.1712606.

Elbeih, S. F. (2015). An overview of integrated remote sensing and GIS for groundwater mapping in Egypt. *Ain Shams Engineering Journal*, *6*(1), 1-15. https://doi.org/10.1016/j.asej.2014.08.008.

Elewa, H. H., & Qaddah, A. A. (2011). Groundwater potentiality mapping in the Sinai Peninsula, Egypt, using remote sensing and GIS-watershed-based

modeling. *Hydrogeology Journal*, *19*(3), 613-628.

Elmahdy, S. I., & Mohamed, M. M. (2014). Groundwater potential modelling using remote sensing and GIS: a case study of the Al Dhaid area, United Arab Emirates. *Geocarto International*, *29*(4), 433-450. https://doi.org/10.1080/10106049.2013.784366.

Falah, F., Ghorbani Nejad, S., Rahmati, O., Daneshfar, M., & Zeinivand, H. (2017). Applicability of generalized additive model in groundwater potential modelling and comparison its performance by bivariate statistical methods. *Geocarto international*, *32*(10), 1069-1089. https://doi.org/10.1080/10106049.2016.1188166.

Falah, F., & Zeinivand, H. (2019). Gis-based groundwater potential mapping in khorramabad in lorestan, Iran, using frequency ratio (fr) and weights of evidence (woe) models. *Water Resources*, *46*(5), 679-692.

Fenta, A. A., Kifle, A., Gebreyohannes, T., & Hailu, G. (2015). Spatial analysis of groundwater potential using remote sensing and GIS-based multi-criteria evaluation in Raya Valley, northern Ethiopia. *Hydrogeology Journal*, *23*(1), 195-206. 10.1007/s10040-014-1198-x.

Fischer, D., Charles, E. G., & Baehr, A. L. (2003). Effects of stormwater infiltration on quality of groundwater beneath retention and detention basins. *Journal of Environmental Engineering*, *129*(5), 464-471. : 10.1061/~ASCE!0733-9372~2003!129:5(464).

Francis, R. A. (2011). The impacts of modern warfare on freshwater ecosystems. *Environmental Management*, *48*(5), 985-999. 10.1007/s00267-011-9746-9.

Ganapuram, S., Kumar, G. V., Krishna, I. M., Kahya, E., & Demirel, M. C. (2009). Mapping

of groundwater potential zones in the Musi basin using remote sensing data and GIS. *Advances in Engineering Software*, *40*(7), 506-518. https://doi.org/10.1016/j.advengsoft.2008.10.001.

Gaur, S., Chahar, B. R., & Graillot, D. (2011). Combined use of groundwater modeling and potential zone analysis for management of groundwater. *International Journal of Applied Earth Observation and Geoinformation*, *13*(1), 127-139. https://doi.org/10.1016/j.jag.2010.09.001.

Gayo, E. M., Latorre, C., Jordan, T. E., Nester, P. L., Estay, S. A., Ojeda, K. F., & Santoro, C. M. (2012). Late Quaternary hydrological and ecological changes in the hyperarid core of the northern Atacama Desert (~ 21 S). *Earth-Science Reviews*, *113*(3-4), 120-140. https://doi.org/10.1016/j.earscirev.2012.04.003.

Ghaffari, M., Chavoshbashi, A. A., Eslami, A., Hatami, H., Pourakbar, M., & Hashemi, M. (2021). Spatial and temporal variation of groundwater quality around a volcanic mountain in northwest of Iran. *Groundwater for Sustainable Development*, *14*, 100627. https://doi.org/10.1016/j.gsd.2021.100627.

Ghorbani Nejad, S., Falah, F., Daneshfar, M., Haghizadeh, A., & Rahmati, O. (2017). Delineation of groundwater potential zones using remote sensing and GIS-based data-driven models. *Geocarto international*, *32*(2), 167-187. https://doi.org/10.1080/10106049.2015.1132481.

Ghosh, P. K., Bandyopadhyay, S., & Jana, N. C. (2016). Mapping of groundwater potential zones in hard rock terrain using geoinformatics: a case of Kumari watershed in western part of West Bengal. *Modeling Earth Systems and Environment*, *2*(1), 1-12. 10.1007/s40808-015-0044-z.

Gnanachandrasamy, G., Zhou, Y., Bagyaraj, M., Venkatramanan, S., Ramkumar, T., & Wang, S. (2018). Remote sensing and GIS based groundwater potential zone mapping in Ariyalur District, Tamil Nadu. *Journal of the Geological Society of*

India, *92*(4), 484-490. 10.1007/s12594-018-1046-z.

Goldstein, B. A., Polley, E. C., & Briggs, F. B. (2011). Random forests for genetic association studies. *Statistical applications in genetics and molecular biology*, *10*(1). https://doi.org/10.2202/1544-6115.1691.

Golkarian, A., Naghibi, S. A., Kalantar, B., & Pradhan, B. (2018). Groundwater potential mapping using C5. 0, random forest, and multivariate adaptive regression spline models in GIS. *Environmental monitoring and assessment*, *190*(3), 1-16. https://doi.org/10.1007/s10661-018-6507-8.

Golkarian, A., & Rahmati, O. (2018). Use of a maximum entropy model to identify the key factors that influence groundwater availability on the Gonabad Plain, Iran. *Environmental Earth Sciences*, *77*(10), 1-20. https://doi.org/10.1007/s12665-018-7551-y.

Golla, V. (2020). Delineation of groundwater potential zones in Sathyavedu area, Chittoor District (Andhra Pradesh), South India, using geospatial technologies. *Modeling Earth Systems and Environment*, *6*(2), 895-905. https://doi.org/10.1007/s40808-020-00726-9.

Gopinath, S., Srinivasamoorthy, K., Saravanan, K., Prakash, R., & Karunanidhi, D. (2019). Characterizing groundwater quality and seawater intrusion in coastal aquifers of Nagapattinam and Karaikal, South India using hydrogeochemistry and modeling techniques. *Human and Ecological Risk Assessment: An International Journal*, *25*(1-2), 314-334. https://doi.org/10.1080/10807039.2019.1578947.

Greer, J. (2008). Safe Yield Estimate for the Beryl-Enterprise Area.

Gulgundi, M. S., & Shetty, A. (2018). Groundwater quality assessment of urban Bengaluru using multivariate statistical techniques. *Applied water science*, *8*(1), 1-15. https://doi.org/10.1007/s13201-018-0684-z.

Gumma, M. K., & Pavelic, P. (2013). Mapping of groundwater potential zones across Ghana using remote sensing, geographic information systems, and spatial modeling. *Environmental monitoring and assessment*, *185*(4), 3561-3579. 10.1007/s10661-012-2810-y.

Gupta, M., & Srivastava, P. K. (2010). Integrating GIS and remote sensing for identification of groundwater potential zones in the hilly terrain of Pavagarh, Gujarat, India. *Water International*, *35*(2), 233-245. https://doi.org/10.1080/02508061003664419.

Guru, B., Seshan, K., & Bera, S. (2017). Frequency ratio model for groundwater potential mapping and its sustainable management in cold desert, India. *Journal of King Saud University-Science*, *29*(3), 333-347. https://doi.org/10.1016/j.jksus.2016.08.003.

Guzman, S. M., Paz, J. O., Tagert, M. L. M., & Mercer, A. E. (2019). Evaluation of seasonally classified inputs for the prediction of daily groundwater levels: NARX networks vs support vector machines. *Environmental Modeling & Assessment*, *24*(2), 223-234. https://doi.org/10.1007/s10666-018-9639-x.

Haridas, V. R., Aravindan, S., & Girish, G. (1998). Remote sensing and its applications for groundwater favourable area identification. *Quarterly Journal of GARC*, *6*(6), 18-22.

Hashim, M., Ahmad, S., Johari, M. A. M., & Pour, A. B. (2013). Automatic lineament extraction in a heavily vegetated region using Landsat Enhanced Thematic Mapper (ETM+) imagery. *Advances in Space Research*, *51*(5), 874-890. https://doi.org/10.1016/j.asr.2012.10.004.

Hou, C., Wen, Y., Liu, X., & Dong, M. (2021). Impacts of regional water shortage information disclosure on public acceptance of recycled water—evidences from China's urban residents. *Journal of Cleaner Production*, *278*, 123965.

https://doi.org/10.1016/j.jclepro.2020.123965.

Hounsinou, S. P. (2020). Assessment of potential seawater intrusion in a coastal aquifer system at Abomey-Calavi, Benin. *Heliyon*, *6*(2), e03173. https://doi.org/10.1016/j.heliyon.2020.e03173.

Howard, G., Charles, K., Pond, K., Brookshaw, A., Hossain, R., & Bartram, J. (2010). Securing 2020 vision for 2030: climate change and ensuring resilience in water and sanitation services. *Journal of water and climate change*, *1*(1), 2-16. https://doi.org/10.2166/wcc.2010.105b.

Hu, H., Jin, Q., & Kavan, P. (2014). A study of heavy metal pollution in China: Current status, pollution-control policies and countermeasures. *Sustainability*, *6*(9), 5820-5838. https://doi.org/10.3390/su6095820.

Ibrahim-Bathis, K., & Ahmed, S. A. (2016). Geospatial technology for delineating groundwater potential zones in Doddahalla watershed of Chitradurga district, India. *The Egyptian Journal of Remote Sensing and Space Science*, *19*(2), 223-234. https://doi.org/10.1016/j.ejrs.2016.06.002.

Jaafari, A., Najafi, A., Pourghasemi, H. R., Rezaeian, J., & Sattarian, A. (2014). GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran. *International Journal of Environmental Science and Technology*, *11*(4), 909-926. 10.1007/s13762-013-0464-0.

Janakarajan, S., & Moench, M. (2006). Are wells a potential threat to farmers' well-being? Case of deteriorating groundwater irrigation in Tamil Nadu. *Economic and Political Weekly*, 3977-3987.

Jenifer, M. A., & Jha, M. K. (2017). Comparison of analytic hierarchy process, catastrophe and entropy techniques for evaluating groundwater prospect of

hard-rock aquifer systems. *Journal of hydrology*, *548*, 605-624. https://doi.org/10.1016/j.jhydrol.2017.03.023.

Jha, M. K., Chowdary, V. M., & Chowdhury, A. (2010). Groundwater assessment in Salboni Block, West Bengal (India) using remote sensing, geographical information system and multi-criteria decision analysis techniques. *Hydrogeology journal*, *18*(7), 1713-1728. 10.1007/s10040-010-0631-z.

Jha, M. K., Shekhar, A., & Jenifer, M. A. (2020). Assessing groundwater quality for drinking water supply using hybrid fuzzy-GIS-based water quality index. *Water Research*, *179*, 115867. https://doi.org/10.1016/j.watres.2020.115867.

Jothibasu, A., & Anbazhagan, S. (2016). Modeling groundwater probability index in Ponnaiyar River basin of South India using analytic hierarchy process. *Modeling Earth Systems and Environment*, *2*(3), 1-14. 10.1007/s40808-016-0174-y.

Kaky, E., Nolan, V., Alatawi, A., & Gilbert, F. (2020). A comparison between Ensemble and MaxEnt species distribution modelling approaches for conservation: A case study with Egyptian medicinal plants. *Ecological Informatics*, *60*, 101150. https://doi.org/10.1016/j.ecoinf.2020.101150.

Kalantar, B., Al-Najjar, H. A., Pradhan, B., Saeidi, V., Halin, A. A., Ueda, N., & Naghibi, S. A. (2019). Optimized conditioning factors using machine learning techniques for groundwater potential mapping. *Water*, *11*(9), 1909. https://doi.org/10.3390/w11091909.

Kallali, H., Anane, M., Jellali, S., & Tarhouni, J. (2007). GIS-based multi-criteria analysis for potential wastewater aquifer recharge sites. *Desalination*, *215*(1-3), 111-119. https://doi.org/10.1016/j.desal.2006.11.016.

Kamali Maskooni, E., Naghibi, S. A., Hashemi, H., & Berndtsson, R. (2020). Application

of advanced machine learning algorithms to assess groundwater potential using remote sensing-derived data. *Remote Sensing*, *12*(17), 2742. https://doi.org/10.3390/rs12172742.

Kawo, N. S., & Karuppannan, S. (2018). Groundwater quality assessment using water quality index and GIS technique in Modjo River Basin, central Ethiopia. *Journal of African Earth Sciences*, *147*, 300-311. https://doi.org/10.1016/j.jafrearsci.2018.06.034.

Kebede, S. (2013). Groundwater potential, recharge, water balance: Vital numbers. In *Groundwater in Ethiopia* (pp. 221-236). Springer, Berlin, Heidelberg. 10.1007/978-3-642-30391-3.

Keesstra, S. D., Geissen, V., Mosse, K., Piiranen, S., Scudiero, E., Leistra, M., & van Schaik, L. (2012). Soil as a filter for groundwater quality. *Current Opinion in Environmental Sustainability*, *4*(5), 507-516. https://doi.org/10.1016/j.cosust.2012.10.007.

Kerry, R., & Oliver, M. A. (2007). Determining the effect of asymmetric data on the variogram. I. Underlying asymmetry. *Computers & geosciences*, *33*(10), 1212-1232. https://doi.org/10.1016/j.cageo.2007.05.008.

Khal, M., Algouti, A., Algouti, A., Akdim, N., Stankevich, S. A., & Menenti, M. (2020). Evaluation of open Digital Elevation Models: estimation of topographic indices relevant to erosion risk in the Wadi M'Goun watershed, Morocco. *AIMS Geosci*, *6*, 231-257. 10.3934/geosci.2020014.

Khatri, N., & Tyagi, S. (2015). Influences of natural and anthropogenic factors on surface and groundwater quality in rural and urban areas. *Frontiers in life science*, *8*(1), 23-39. https://doi.org/10.1080/21553769.2014.933716.

Khoshtinat, S., Aminnejad, B., Hassanzadeh, Y., & Ahmadi, H. (2019). Groundwater

potential assessment of the Sero plain using bivariate models of the frequency ratio, Shannon entropy and evidential belief function. *Journal of Earth System Science*, *128*(6), 1-16. https://doi.org/10.1007/s12040-019-1155-0.

Khosravi, K., Panahi, M., & Tien Bui, D. (2018). Spatial prediction of groundwater spring potential mapping based on an adaptive neuro-fuzzy inference system and metaheuristic optimization. Hydrology and Earth System Sciences, 22(9), 4771-4792. https://doi.org/10.5194/hess-22-4771-2018.

Kim, J. C., Jung, H. S., & Lee, S. (2019). Spatial mapping of the groundwater potential of the geum river basin using ensemble models based on remote sensing images. *Remote Sensing*, *11*(19), 2285. https://doi.org/10.3390/rs11192285.

Klongvessa, P., Lu, M., & Chotpantarat, S. (2018). Variations of characteristics of consecutive rainfall days over northern Thailand. *Theoretical and Applied Climatology*, *133*(3), 737-749. 10.1007/s00704-017-2208-4.

Kohavi, R. (1995, August). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai* (Vol. 14, No. 2, pp. 1137-1145).

Konkul, J., Rojborwornwittaya, W., & Chotpantarat, S. (2014). Hydrogeologic characteristics and groundwater potentiality mapping using potential surface analysis in the Huay Sai area, Phetchaburi province, Thailand. *Geosciences Journal*, *18*(1), 89-103. 10.1007/s12303-013-0047-6.

Kordestani, M. D., Naghibi, S. A., Hashemi, H., Ahmadi, K., Kalantar, B., & Pradhan, B. (2019). Groundwater potential mapping using a novel data-mining ensemble model. *Hydrogeology journal*, *27*(1), 211-224. https://doi.org/10.1007/s10040-018-1848-5.

Kresic, N. (2010). Types and classifications of springs. In *Groundwater hydrology of springs* (pp. 31-85). Butterworth-Heinemann. https://doi.org/10.1016/B978-1-

85617-502-9.00002-5.

Kumar, A., & Krishna, A. P. (2018). Assessment of groundwater potential zones in coal mining impacted hard-rock terrain of India by integrating geospatial and analytic hierarchy process (AHP) approach. *Geocarto International*, *33*(2), 105-129. https://doi.org/10.1080/10106049.2016.1232314.

Kumar, D. (2018). Evolving Differential evolution method with random forest for prediction of Air Pollution. Procedia computer science, 132, 824-833. https://doi.org/10.1016/j.procs.2018.05.094.

Kumar, P. J., & Augustine, C. M. (2022). Entropy-weighted water quality index (EWQI) modeling of groundwater quality and spatial mapping in Uppar Odai Sub-Basin, South India. *Modeling earth systems and environment*, *8*(1), 911-924. https://doi.org/10.1007/s40808-021-01132-5.

Kumar, P., Herath, S., Avtar, R., & Takeuchi, K. (2016). Mapping of groundwater potential zones in Killinochi area, Sri Lanka, using GIS and remote sensing techniques. *Sustainable Water Resources Management*, *2*(4), 419-430. 10.1007/s40899-016-0072-5.

Kumar, T., Gautam, A. K., & Kumar, T. (2014). Appraising the accuracy of GIS-based multi-criteria decision making technique for delineation of groundwater potential zones. *Water resources management*, *28*(13), 4449-4466. 10.1007/s11269-014-0663-6.

Lattman, L. H., & Parizek, R. R. (1964). Relationship between fracture traces and the occurrence of ground water in carbonate rocks. *Journal of hydrology*, *2*(2), 73-91. https://doi.org/10.1016/0022-1694(64)90019-8.

Lee, K. Y., Kim, K. H., Kang, J. J., Choi, S. J., Im, Y. S., Lee, Y. D., & Lim, Y. S. (2017). Comparison and analysis of linear regression & artificial neural

network. *International Journal of Applied Engineering Research*, *12*(20), 9820-9825.

Lee, S., Hong, S. M., & Jung, H. S. (2018). GIS-based groundwater potential mapping using artificial neural network and support vector machine models: the case of Boryeong city in Korea. *Geocarto international*, *33*(8), 847-861. https://doi.org/10.1080/10106049.2017.1303091.

Lee, S., Hyun, Y., & Lee, M. J. (2019). Groundwater potential mapping using data mining models of big data analysis in Goyang-si, South Korea. *Sustainability*, *11*(6), 1678. https://doi.org/10.3390/su11061678.

Lee, S., Kim, Y. S., & Oh, H. J. (2012). Application of a weights-of-evidence method and GIS to regional groundwater productivity potential mapping. *Journal of Environmental Management*, *96*(1), 91-105. https://doi.org/10.1016/j.jenvman.2011.09.016.

Lee, S., & Pradhan, B. (2006). Probabilistic landslide hazards and risk mapping on Penang Island, Malaysia. *Journal of Earth System Science*, *115*(6), 661-672.

Li, P., He, S., Yang, N., & Xiang, G. (2018). Groundwater quality assessment for domestic and agricultural purposes in Yan'an City, northwest China: implications to sustainable groundwater quality management on the Loess Plateau. *Environmental Earth Sciences*, *77*(23), 1-16. https://doi.org/10.1007/s12665-018-7968-3.

Li, P., Tian, R., Xue, C., & Wu, J. (2017). Progress, opportunities, and key fields for groundwater quality research under the impacts of human activities in China with a special focus on western China. *Environmental Science and Pollution Research*, *24*(15), 13224-13234. 10.1007/s11356-017-8753-7.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R*

news, *2*(3), 18-22.

Lipczynska-Kochany, E. (2018). Effect of climate change on humic substances and associated impacts on the quality of surface water and groundwater: A review. *Science of the total environment*, *640*, 1548-1565. https://doi.org/10.1016/j.scitotenv.2018.05.376.

Lone, M. S., Nagaraju, D., Mahadavesamy, G., & Siddalingamurthy, S. (2013). Applications of GIS and remote sensing to delineate artificial recharge zones (DARZ) of groundwater in HD Kote taluk, Mysore district, Karnataka, India. *Int J Remote Sens Geosci*, *2*(3), 92-97.

MacDonald, A. M., & Davies, J. (2000). A brief review of groundwater for rural water supply in sub-Saharan Africa. https://nora.nerc.ac.uk/id/eprint/501047.

Machiwal, D., Jha, M. K., & Mal, B. C. (2011). Assessment of groundwater potential in a semi-arid region of India using remote sensing, GIS and MCDM techniques. *Water resources management*, *25*(5), 1359-1386. 10.1007/s11269-010-9749-y.

Magesh, N. S., Chandrasekar, N., & Soundranayagam, J. P. (2012). Delineation of groundwater potential zones in Theni district, Tamil Nadu, using remote sensing, GIS and MIF techniques. *Geoscience frontiers*, *3*(2), 189-196. https://doi.org/10.1016/j.gsf.2011.10.007.

Maggirwar, B.C.,& Umrikar, B.N. (2011). Influence of various factors on the fluctuation of groundwater level in hard rock terrain and its importance in the assessment of groundwater. *Journal of Geology and Mining Research*, *3*(11), 305-317.

Malczewski, J. (1999). *GIS and multicriteria decision analysis*. John Wiley & Sons.

Mallick, J., Naikoo, M. W., Talukdar, S., Ahmed, I. A., Rahman, A., Islam, A. R. M. T., Pal, S., Ghose, S., & Shashtri, S. (2021). Developing groundwater potentiality models

by coupling ensemble machine learning algorithms and statistical techniques for sustainable groundwater management. *Geocarto International*, 1-27. https://doi.org/10.1080/10106049.2021.1987535.

Mallick, J., Singh, C. K., Al‑Wadi, H., Ahmed, M., Rahman, A., Shashtri, S., & Mukherjee, S. (2015). Geospatial and geostatistical approach for groundwater potential zone delineation. *Hydrological Processes*, *29*(3), 395-418. https://doi.org/10.1002/hyp.10153.

Manap, M. A., Nampak, H., Pradhan, B., Lee, S., Sulaiman, W. N. A., & Ramli, M. F. (2014). Application of probabilistic-based frequency ratio model in groundwater potential mapping using remote sensing data and GIS. *Arabian Journal of Geosciences*, *7*(2), 711-724. 10.1007/s12517-012-0795-z.

Mandal, U., Sahoo, S., Munusamy, S. B., Dhar, A., Panda, S. N., Kar, A., & Mishra, P. K. (2016). Delineation of groundwater potential zones of coastal groundwater basin using multi-criteria decision making technique. *Water Resources Management*, *30*(12), 4293-4310. 10.1007/s11269-016-1421-8.

Marendra, S. M. P., & Tangahu, B. V. (2020, May). Study of Leachate Penetration in Shallow Groundwater Around Jabon Landfill Sidoarjo. In *IOP Conference Series: Earth and Environmental Science* (Vol. 506, No. 1, p. 012034). IOP Publishing. 10.1088/1755-1315/506/1/012034.

Martha, T. R., Ghosh, D., Kumar, K. V., Lesslie, A., & Kumar, M. V. (2013). Geospatial technologies for national geomorphology and lineament mapping project—a case study of Goa state. *Journal of the Indian Society of Remote Sensing*, *41*(4), 905-920. 10.1007/s12524-012-0260-1.

Martínez‑Santos, P., & Renard, P. (2020). Mapping groundwater potential through an ensemble of big data methods. *Groundwater*, *58*(4), 583-597.

https://doi.org/10.1111/gwat.12939.

McDonald, R. I., Green, P., Balk, D., Fekete, B. M., Revenga, C., Todd, M., & Montgomery, M. (2011). Urban growth, climate change, and freshwater availability. *Proceedings of the National Academy of Sciences*, *108*(15), 6312-6317. https://doi.org/10.1073/pnas.1011615108.

McLay, C. D. A., Dragten, R., Sparling, G., & Selvarajah, N. (2001). Predicting groundwater nitrate concentrations in a region of mixed agricultural land use: a comparison of three approaches. *Environmental pollution*, *115*(2), 191-204. https://doi.org/10.1016/S0269-7491(01)00111-7.

Miraki, S., Zanganeh, S. H., Chapi, K., Singh, V. P., Shirzadi, A., Shahabi, H., & Pham, B. T. (2019). Mapping groundwater potential using a novel hybrid intelligence approach. *Water resources management*, *33*(1), 281-302. https://doi.org/10.1007/s11269-018-2102-6.

Mogaji, K. A., Lim, H. S., & Abdullah, K. (2015). Regional prediction of groundwater potential mapping in a multifaceted geology terrain using GIS-based Dempster–Shafer model. *Arabian Journal of Geosciences*, *8*(5), 3235-3258. 10.1007/s12517-014-1391-1.

Moglen, G. E., Eltahir, E. A., & Bras, R. L. (1998). On the sensitivity of drainage density to climate change. *Water resources research*, *34*(4), 855-862. https://doi.org/10.1029/97WR02709.

Mohan, C., Western, A. W., Wei, Y., & Saft, M. (2018). Predicting groundwater recharge for varying land cover and climate conditions–a global meta-study. *Hydrology and Earth System Sciences*, *22*(5), 2689-2703. https://doi.org/10.5194/hess-22-2689-2018, 2018.

Mohankumar, K., Hariharan, V., & Rao, N. P. (2016). Heavy metal contamination in

groundwater around industrial estate vs residential areas in Coimbatore, India. *Journal of clinical and diagnostic research: JCDR*, *10*(4), BC05. 10.7860/JCDR/2016/15943.7527.

Mollinedo, J., Schumacher, T. E., & Chintala, R. (2015). Influence of feedstocks and pyrolysis on biochar's capacity to modify soil water retention characteristics. *Journal of Analytical and Applied Pyrolysis*, *114*, 100-108. https://doi.org/10.1016/j.jaap.2015.05.006.

Monteiro, A. L., de Freitas Souza, M., Lins, H. A., da Silva Teófilo, T. M., Júnior, A. P. B., Silva, D. V., & Mendonça, V. (2021). A new alternative to determine weed control in agricultural systems based on artificial neural networks (ANNs). *Field Crops Research*, *263*, 108075. https://doi.org/10.1016/j.fcr.2021.108075.

Moore, G. K., & Waltz, F. A. (1983). Objective procedures for lineament enhancement and extraction. *Photogrammetric Engineering and remote sensing*, *49*(5), 641-647.

Mosavi, A., Sajedi Hosseini, F., Choubin, B., Goodarzi, M., Dineva, A. A., & Rafiei Sardooi, E. (2021). Ensemble boosting and bagging based machine learning models for groundwater potential prediction. *Water Resources Management*, *35*(1), 23-37. https://doi.org/10.1007/s11269-020-02704-3.

Mousavi, S. M., Golkarian, A., Naghibi, S. A., Kalantar, B., & Pradhan, B. (2017). GIS-based groundwater spring potential mapping using data mining boosted regression tree and probabilistic frequency ratio models in Iran. *Aims Geosci*, *3*(1), 91-115. 10.3934/geosci.2017.1.91.

Mousazade, M., Ghanbarian, G., Pourghasemi, H. R., Safaeian, R., & Cerdà, A. (2019). Maxent data mining technique and its comparison with a bivariate statistical model for predicting the potential distribution of Astragalus Fasciculifolius Boiss. in Fars, Iran. *Sustainability, 11*(12), 3452.

https://doi.org/10.3390/su11123452.

Muavhi, N., Thamaga, K. H., & Mutoti, M. I. (2021). Mapping groundwater potential zones using relative frequency ratio, analytic hierarchy process and their hybrid models: case of Nzhelele-Makhado area in South Africa. *Geocarto International*, 1-20. https://doi.org/10.1080/10106049.2021.1936212.

Muchingami, I., Hlatywayo, D. J., Nel, J. M., & Chuma, C. (2012). Electrical resistivity survey for groundwater investigations and shallow subsurface evaluation of the basaltic-greenstone formation of the urban Bulawayo aquifer. *Physics and Chemistry of the Earth, Parts A/B/C*, *50*, 44-51. https://doi.org/10.1016/j.pce.2012.08.014.

Mukherjee, P., Singh, C. K., & Mukherjee, S. (2012). Delineation of groundwater potential zones in arid region of India—a remote sensing and GIS approach. *Water resources management*, *26*(9), 2643-2672. 10.1007/s11269-012-0038-9.

Mumtaz, R., Baig, S., Kazmi, S. S. A., Ahmad, F., Fatima, I., & Ghauri, B. (2019). Delineation of groundwater prospective resources by exploiting geo-spatial decision-making techniques for the Kingdom of Saudi Arabia. *Neural Computing and Applications*, *31*(9), 5379-5399. https://doi.org/10.1007/s00521-018-3370-z.

Murmu, P., Kumar, M., Lal, D., Sonker, I., & Singh, S. K. (2019). Delineation of groundwater potential zones using geospatial techniques and analytical hierarchy process in Dumka district, Jharkhand, India. *Groundwater for Sustainable Development*, *9*, 100239. https://doi.org/10.1016/j.gsd.2019.100239.

Mwaniki, M. W., Moeller, M. S., & Schellmann, G. (2015). A comparison of Landsat 8 (OLI) and Landsat 7 (ETM+) in mapping geology and visualising lineaments: A case study of central region Kenya. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.

10.5194/isprsarchives-XL-7-W3-897-2015.

Nafouanti, M. B., Li, J., Mustapha, N. A., Uwamungu, P., & Dalal, A. A. (2021). Prediction on the fluoride contamination in groundwater at the Datong Basin, Northern China: Comparison of random forest, logistic regression and artificial neural network. *Applied Geochemistry*, *132*, 105054. https://doi.org/10.1016/j.apgeochem.2021.105054.

Nag, S. K., & Ghosh, P. (2013). Delineation of groundwater potential zone in Chhatna Block, Bankura District, West Bengal, India using remote sensing and GIS techniques. *Environmental earth sciences*, *70*(5), 2115-2127. 10.1007/s12665-012-1713-0.

Naghibi, S. A., Ahmadi, K., & Daneshi, A. (2017a). Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. *Water Resources Management*, *31*(9), 2761-2775. 10.1007/s11269-017-1660-3.

Naghibi, S. A., Dolatkordestani, M., Rezaei, A., Amouzegari, P., Heravi, M. T., Kalantar, B., & Pradhan, B. (2019). Application of rotation forest with decision trees as base classifier and a novel ensemble model in spatial modeling of groundwater potential. Environmental monitoring and assessment, 191(4), 1-20. https://doi.org/10.1007/s10661-019-7362-y.

Naghibi, S. A., Moghaddam, D. D., Kalantar, B., Pradhan, B., & Kisi, O. (2017b). A comparative assessment of GIS-based data mining models and a novel ensemble model in groundwater well potential mapping. *Journal of Hydrology*, *548*, 471-483. https://doi.org/10.1016/j.jhydrol.2017.03.020.

Naghibi, S. A., & Moradi Dashtpagerdi, M. (2017). Evaluation of four supervised learning methods for groundwater spring potential mapping in Khalkhal region (Iran) using GIS-based features. *Hydrogeology journal*, *25*(1), 169-189. 10.1007/s10040-

016-1466-z.

Naghibi, S. A., & Pourghasemi, H. R. (2015). 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*, *29*(14), 5217-5236. 10.1007/s11269-015-1114-8.

Naghibi, S. A., Pourghasemi, H. R., & Abbaspour, K. (2018). A comparison between ten advanced and soft computing models for groundwater qanat potential assessment in Iran using R and GIS. *Theoretical and applied climatology*, *131*(3), 967-984. 10.1007/s00704-016-2022-4.

Naghibi, S. A., Pourghasemi, H. R., & Dixon, B. (2016). 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*, *188*(1), 1-27. 10.1007/s10661-015-5049-6.

Nampak, H., Pradhan, B., & Abd Manap, M. (2014). Application of GIS based data driven evidential belief function model to predict groundwater potential zonation. *Journal of Hydrology*, *513*, 283-300. https://doi.org/10.1016/j.jhydrol.2014.02.053.

Nanda, S., Annadurai, R., & Barik, K. K. (2017). Geospatial decipherment of groundwater potential of Kattankolathur block of Tamil Nadu using MCDM techniques. *Remote Sensing Applications: Society and Environment*, *8*, 240-250. https://doi.org/10.1016/j.rsase.2017.10.002.

Nasir, M. J., Khan, S., Zahid, H., & Khan, A. (2018). Delineation of groundwater potential zones using GIS and multi influence factor (MIF) techniques: a study of district Swat, Khyber Pakhtunkhwa, Pakistan. *Environmental Earth Sciences*, *77*(10), 1-11. https://doi.org/10.1007/s12665-018-7522-3.

Nguyen, D., Gupta, S., Rana, S., Shilton, A., & Venkatesh, S. (2020a). Bayesian optimization for categorical and category-specific continuous inputs. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 04, pp. 5256-5263). https://doi.org/10.1609/aaai.v34i04.5971.

Nguyen, P. T., Ha, D. H., Avand, M., Jaafari, A., Nguyen, H. D., Al-Ansari, N., Phong, T. V., Sharma, R., Kumar, R., Le, H. V., Ho, L. S., Prakash, I., & Pham, B. T. (2020b). Soft computing ensemble models based on logistic regression for groundwater potential mapping. *Applied Sciences*, *10*(7), 2469. https://doi.org/10.3390/app10072469.

Nguyen, P. T., Ha, D. H., Jaafari, A., Nguyen, H. D., Van Phong, T., Al-Ansari, N., Le, H.V., & Pham, B. T. (2020c). Groundwater potential mapping combining artificial neural network and real AdaBoost ensemble technique: the DakNong province case-study, Vietnam. *International journal of environmental research and public health*, *17*(7), 2473. https://doi.org/10.3390/ijerph17072473.

Noori, R., Karbassi, A., & Sabahi, M. S. (2010). Evaluation of PCA and Gamma test techniques on ANN operation for weekly solid waste prediction. *Journal of Environmental Management*, *91*(3), 767-771. https://doi.org/10.1016/j.jenvman.2009.10.007.

O'leary, D. W., Friedman, J. D., & Pohn, H. A. (1976). Lineament, linear, lineation: some proposed new standards for old terms. *Geological Society of America Bulletin*, *87*(10), 1463-1469. https://doi.org/10.1130/0016-7606(1976)87<1463:LLLSPN>2.0.CO;2.

Odeh, T., Mohammad, A. H., Hussein, H., Ismail, M., & Almomani, T. (2019). Over-pumping of groundwater in Irbid governorate, northern Jordan: a conceptual model to analyze the effects of urbanization and agricultural activities on groundwater levels and salinity. *Environmental Earth Sciences*, *78*(1), 1-12.

https://doi.org/10.1007/s12665-018-8031-0.

Oh, H. J., Kim, Y. S., Choi, J. K., Park, E., & Lee, S. (2011). GIS mapping of regional probabilistic groundwater potential in the area of Pohang City, Korea. *Journal of Hydrology*, *399*(3-4), 158-172. https://doi.org/10.1016/j.jhydrol.2010.12.027.

Oikonomidis, D., Dimogianni, S., Kazakis, N., & Voudouris, K. (2015). A GIS/remote sensing-based methodology for groundwater potentiality assessment in Tirnavos area, Greece. *Journal of Hydrology*, *525*, 197-208. https://doi.org/10.1016/j.jhydrol.2015.03.056.

Opitz, D., & Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of artificial intelligence research*, *11*, 169-198. https://doi.org/10.1613/jair.614.

Owolabi, S. T., Madi, K., Kalumba, A. M., & Orimoloye, I. R. (2020). A groundwater potential zone mapping approach for semi-arid environments using remote sensing (RS), geographic information system (GIS), and analytical hierarchical process (AHP) techniques: a case study of Buffalo catchment, Eastern Cape, South Africa. *Arabian journal of geosciences*, *13*(22), 1-17. https://doi.org/10.1007/s12517-020-06166-0.

Owor, M., Taylor, R. G., Tindimugaya, C., & Mwesigwa, D. (2009). Rainfall intensity and groundwater recharge: empirical evidence from the Upper Nile Basin. *Environmental Research Letters*, *4*(3), 035009. 10.1088/1748-9326/4/3/035009.

Ozdemir, A. (2011a). Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the Sultan Mountains (Aksehir, Turkey). *Journal of Hydrology*, *405*(1-2), 123-136. https://doi.org/10.1016/j.jhydrol.2011.05.015.

Ozdemir, A. (2011b). GIS-based groundwater spring potential mapping in the Sultan

Mountains (Konya, Turkey) using frequency ratio, weights of evidence and logistic regression methods and their comparison. *Journal of hydrology*, *411*(3-4), 290-308. https://doi.org/10.1016/j.jhydrol.2011.10.010.

Pailoplee, S., Sugiyama, Y., & Charusiri, P. (2009). Deterministic and probabilistic seismic hazard analyses in Thailand and adjacent areas using active fault data. *Earth, planets and space*, *61*(12), 1313-1325.

Panahi, M., Sadhasivam, N., Pourghasemi, H. R., Rezaie, F., & Lee, S. (2020). Spatial prediction of groundwater potential mapping based on convolutional neural network (CNN) and support vector regression (SVR). *Journal of Hydrology*, *588*, 125033. https://doi.org/10.1016/j.jhydrol.2020.125033.

Papaioannou, G., Vasiliades, L., & Loukas, A. (2015). Multi-criteria analysis framework for potential flood prone areas mapping. *Water resources management*, *29*(2), 399-418. 10.1007/s11269-014-0817-6.

Patra, S., Mishra, P., & Mahapatra, S. C. (2018). Delineation of groundwater potential zone for sustainable development: A case study from Ganga Alluvial Plain covering Hooghly district of India using remote sensing, geographic information system and analytic hierarchy process. *Journal of Cleaner Production*, *172*, 2485-2502. https://doi.org/10.1016/j.jclepro.2017.11.161.

Peiyue, L. I., Jianhua, W. U., & Hui, Q. (2010). Groundwater quality assessment based on entropy weighted osculating value method. *International Journal of Environmental Sciences*, *1*(4), 621-630.

Pham, B. T., Jaafari, A., Prakash, I., Singh, S. K., Quoc, N. K., & Bui, D. T. (2019). Hybrid computational intelligence models for groundwater potential mapping. *Catena*, *182*, 104101. https://doi.org/10.1016/j.catena.2019.104101.

Pholkern, K., Saraphirom, P., & Srisuk, K. (2018). Potential impact of climate change on

groundwater resources in the Central Huai Luang Basin, Northeast Thailand. *Science of the Total Environment*, *633*, 1518-1535. https://doi.org/10.1016/j.scitotenv.2018.03.300.

Pinto, D., Shrestha, S., Babel, M. S., & Ninsawat, S. (2017). Delineation of groundwater potential zones in the Comoro watershed, Timor Leste using GIS, remote sensing and analytic hierarchy process (AHP) technique. *Applied Water Science*, *7*(1), 503-519. 10.1007/s13201-015-0270-6.

Pourghasemi, H. R., & Beheshtirad, M. (2015). Assessment of a data-driven evidential belief function model and GIS for groundwater potential mapping in the Koohrang Watershed, Iran. *Geocarto International*, *30*(6), 662-685. https://doi.org/10.1080/10106049.2014.966161.

Pradhan, A. M. S., Kim, Y. T., Shrestha, S., Huynh, T. C., & Nguyen, B. P. (2020). Application of deep neural network to capture groundwater potential zone in mountainous terrain, Nepal Himalaya. *Environmental Science and Pollution Research*, *28*(15), 18501-18517. https://doi.org/10.1007/s11356-020-10646-x.

Pradhan, B. (2010). Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches. *Journal of the Indian Society of Remote Sensing*, *38*(2), 301-320.

Prasad, P., Loveson, V. J., Kotha, M., & Yadav, R. (2020). Application of machine learning techniques in groundwater potential mapping along the west coast of India. *GIScience & Remote Sensing*, *57*(6), 735-752. https://doi.org/10.1080/15481603.2020.1794104.

Preeja, K. R., Joseph, S., Thomas, J., & Vijith, H. (2011). Identification of groundwater potential zones of a tropical river basin (Kerala, India) using remote sensing and GIS techniques. *Journal of the Indian Society of Remote Sensing*, *39*(1), 83-94. 10.1007/s12524-011-0075-5.

Putthividhya, A., & Pipitsombat, R. (2018, April). Long-Term Assessment and Modeling of Agricultural Nitrate in Groundwater Systems of Thailand. In *EGU General Assembly Conference Abstracts* (p. 2972).

Rabeiy, R. E. (2018). Assessment and modeling of groundwater quality using WQI and GIS in Upper Egypt area. *Environmental Science and Pollution Research*, *25*(31), 30808-30817. 10.1007/s11356-017-8617-1.

Raheja, H., Goel, A., & Pal, M. (2022). Prediction of groundwater quality indices using machine learning algorithms. *Water Practice & Technology*, *17*(1), 336-351. https://doi.org/10.2166/wpt.2021.120.

Rahmati, O., & Melesse, A. M. (2016). Application of Dempster–Shafer theory, spatial analysis and remote sensing for groundwater potentiality and nitrate pollution analysis in the semi-arid region of Khuzestan, Iran. *Science of the Total Environment*, *568*, 1110-1123. https://doi.org/10.1016/j.scitotenv.2016.06.176.

Rahmati, O., Naghibi, S. A., Shahabi, H., Bui, D. T., Pradhan, B., Azareh, A., Rafiei-Sardooi, E., Samani, A. N., & Melesse, A. M. (2018). Groundwater spring potential modelling: Comprising the capability and robustness of three different modeling approaches. *Journal of hydrology*, *565*, 248-261. https://doi.org/10.1016/j.jhydrol.2018.08.027.

Rahmati, O., Pourghasemi, H. R., & Melesse, A. M. (2016). Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran Region, Iran. *Catena*, *137*, 360-372. https://doi.org/10.1016/j.catena.2015.10.010.

Rahmati, O., Nazari Samani, A., Mahdavi, M., Pourghasemi, H. R., & Zeinivand, H. (2015). Groundwater potential mapping at Kurdistan region of Iran using analytic hierarchy process and GIS. *Arabian Journal of Geosciences*, *8*(9), 7059-7071. 10.1007/s12517-014-1668-4.

Rajaee, T., Ebrahimi, H., & Nourani, V. (2019). A review of the artificial intelligence methods in groundwater level modeling. *Journal of hydrology*, *572*, 336-351. https://doi.org/10.1016/j.jhydrol.2018.12.037.

Rajasekhar, M., Raju, G. S., Sreenivasulu, Y., & Raju, R. S. (2019). Delineation of groundwater potential zones in semi-arid region of Jilledubanderu river basin, Anantapur District, Andhra Pradesh, India using fuzzy logic, AHP and integrated fuzzy-AHP approaches. *HydroResearch*, *2*, 97-108. https://doi.org/10.1016/j.hydres.2019.11.006.

Rajaveni, S. P., Brindha, K., & Elango, L. (2017). Geological and geomorphological controls on groundwater occurrence in a hard rock region. *Applied water science*, *7*(3), 1377-1389. 10.1007/s13201-015-0327-6.

Ratchawang, S., Charoenrojying, P., & Chotpantarat, S. (2022). Assessment of atrazine migration in soil and groundwater using nitrate as an indicator in intensively cultivated sugarcane field, Suphan Buri Province, Thailand. *Frontiers in Earth Science*, 1028. 10.3389/feart.2022.855599.

Razandi, Y., Pourghasemi, H. R., Neisani, N. S., & Rahmati, O. (2015). Application of analytical hierarchy process, frequency ratio, and certainty factor models for groundwater potential mapping using GIS. *Earth Science Informatics*, *8*(4), 867-883. 10.1007/s12145-015-0220-8.

Reddy, S., & Dávalos, L. M. (2003). Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography*, *30*(11), 1719-1727. https://doi.org/10.1046/j.1365-2699.2003.00946.x.

Rodell, M., Chen, J., Kato, H., Famiglietti, J. S., Nigro, J., & Wilson, C. R. (2007). Estimating groundwater storage changes in the Mississippi River basin (USA) using GRACE. *Hydrogeology Journal*, *15*(1), 159-166. 10.1007/s10040-006-0103-7.

Rodriguez-Galiano, V., Mendes, M. P., Garcia-Soldado, M. J., Chica-Olmo, M., & Ribeiro, L. (2014). Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain). *Science of the Total Environment*, *476*, 189-206. https://doi.org/10.1016/j.scitotenv.2014.01.001.

Saaty, T. L. (2001). *Decision making for leaders: the analytic hierarchy process for decisions in a complex world*. RWS publications.

Sachdeva, S., & Kumar, B. (2021). Comparison of gradient boosted decision trees and random forest for groundwater potential mapping in Dholpur (Rajasthan), India. *Stochastic Environmental Research and Risk Assessment*, *35*(2), 287-306. https://doi.org/10.1007/s00477-020-01891-0.

Sadat-Noori, S. M., Ebrahimi, K., & Liaghat, A. M. (2014). Groundwater quality assessment using the Water Quality Index and GIS in Saveh-Nobaran aquifer, Iran. *Environmental Earth Sciences*, *71*(9), 3827-3843. 10.1007/s12665-013-2770-8.

Saha, S., Reza, A. H. M., & Roy, M. K. (2019). Hydrochemical evaluation of groundwater quality of the Tista floodplain, Rangpur, Bangladesh. *Applied Water Science*, *9*(8), 1-12. https://doi.org/10.1007/s13201-019-1085-7.

Saha, S. (2017). Groundwater potential mapping using analytical hierarchical process: a study on Md. Bazar Block of Birbhum District, West Bengal. *Spatial Information Research*, *25*(4), 615-626. 10.1007/s41324-017-0127-1.

Sahoo, S., Russo, T. A., Elliott, J., & Foster, I. (2017). Machine learning algorithms for modeling groundwater level changes in agricultural regions of the US. *Water Resources Research*, *53*(5), 3878-3895. https://doi.org/10.1002/2016WR019933.

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. *Science of the total environment*, *644*, 954-962. https://doi.org/10.1016/j.scitotenv.2018.07.054.

Şakar, B. G. B. E. (2018). Variable Importance Analysis in Default Prediction using Machine Learning Techniques.

Sameen, M. I., Pradhan, B., & Lee, S. (2019). Self-learning random forests model for mapping groundwater yield in data-scarce areas. *Natural Resources Research*, *28*(3), 757-775. https://doi.org/10.1007/s11053-018-9416-1.

Santhi, C., Arnold, J. G., Williams, J. R., Dugas, W. A., Srinivasan, R., & Hauck, L. M. (2001). Validation of the swat model on a large rwer basin with point and nonpoint sources 1. *JAWRA Journal of the American Water Resources Association*, *37*(5), 1169-1188. https://doi.org/10.1111/j.1752-1688.2001.tb03630.x.

Saranya, T., & Saravanan, S. (2020). Groundwater potential zone mapping using analytical hierarchy process (AHP) and GIS for Kancheepuram District, Tamilnadu, India. *Modeling Earth Systems and Environment*, *6*(2), 1105-1122. https://doi.org/10.1007/s40808-020-00744-7.

Saravanan, S., Jennifer, J. J., Singh, L., Thiyagarajan, S., & Sankaralingam, S. (2021). Impact of land-use change on soil erosion in the Coonoor Watershed, Nilgiris Mountain Range, Tamil Nadu, India. *Arabian Journal of Geosciences*, *14*(5), 1-15. https://doi.org/10.1007/s12517-021-06817-w.

Schoppa, L., Disse, M., & Bachmair, S. (2020). Evaluating the performance of random forest for large-scale flood discharge simulation. *Journal of Hydrology*, *590*, 125531. https://doi.org/10.1016/j.jhydrol.2020.125531.

Selvam, S., Dar, F. A., Magesh, N. S., Venkatramanan, S., & Chung, S. Y. (2016).

Application of remote sensing and GIS for delineating groundwater recharge potential zones of Kovilpatti Municipality, Tamil Nadu using IF technique. *Earth Science Informatics*, *9*(2), 137-150. 10.1007/s12145-015-0242-2.

Selvam, S., Magesh, N. S., Chidambaram, S., Rajamanickam, M., & Sashikkumar, M. C. (2015). A GIS based identification of groundwater recharge potential zones using RS and IF technique: a case study in Ottapidaram taluk, Tuticorin district, Tamil Nadu. *Environmental earth sciences*, *73*(7), 3785-3799. 10.1007/s12665-014-3664-0.

Shah, T. (2005). Groundwater and human development: challenges and opportunities in livelihoods and environment. *Water Science and Technology*, *51*(8), 27-37. https://doi.org/10.2166/wst.2005.0217.

Shah, T., Molden, D., Sakthivadivel, R., & Seckler, D. (2001). Global groundwater situation: Opportunities and challenges. *Economic and Political Weekly*, 4142-4150. https://www.jstor.org/stable/4411304.

Shao, Z., Huq, M. E., Cai, B., Altan, O., & Li, Y. (2020). Integrated remote sensing and GIS approach using Fuzzy-AHP to delineate and identify groundwater potential zones in semi-arid Shanxi Province, China. *Environmental Modelling & Software*, *134*, 104868. https://doi.org/10.1016/j.envsoft.2020.104868.

Shekhar, S., & Pandey, A. C. (2015). Delineation of groundwater potential zone in hard rock terrain of India using remote sensing, geographical information system (GIS) and analytic hierarchy process (AHP) techniques. *Geocarto International*, *30*(4), 402-421. https://doi.org/10.1080/10106049.2014.894584.

Shen, S. L., Ma, L., Xu, Y. S., & Yin, Z. Y. (2013). Interpretation of increased deformation rate in aquifer IV due to groundwater pumping in Shanghai. *Canadian Geotechnical Journal*, *50*(11), 1129-1142. https://doi.org/10.1139/cgj-2013-0042.

Singh, C. K., Shashtri, S., Singh, A., & Mukherjee, S. (2011). Quantitative modeling of groundwater in Satluj River basin of Rupnagar district of Punjab using remote sensing and geographic information system. *Environmental Earth Sciences*, *62*(4), 871-881. 10.1007/s12665-010-0574-7.

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. *Ecological Indicators*, *91*, 24-37. https://doi.org/10.1016/j.ecolind.2018.03.070.

Singh, L. K., Jha, M. K., & Chowdary, V. M. (2022). Application of catastrophe theory to spatial analysis of groundwater potential in a sub-humid tropical region: a hybrid approach. *Geocarto International*, *37*(3), 700-719. https://doi.org/10.1080/10106049.2020.1737970.

Singh, P., Thakur, J. K., & Kumar, S. (2013). Delineating groundwater potential zones in a hard-rock terrain using geospatial tool. *Hydrological Sciences Journal*, *58*(1), 213-223. https://doi.org/10.1080/02626667.2012.745644.

Singha, S., Pasupuleti, S., Singha, S. S., & Kumar, S. (2020). Effectiveness of groundwater heavy metal pollution indices studies by deep-learning. *Journal of Contaminant Hydrology*, *235*, 103718. https://doi.org/10.1016/j.jconhyd.2020.103718.

Singha, S., Pasupuleti, S., Singha, S. S., Singh, R., & Kumar, S. (2021). Prediction of groundwater quality using efficient machine learning technique. *Chemosphere*, *276*, 130265. https://doi.org/10.1016/j.chemosphere.2021.130265.

Soltani Mohammadi, A., Sayadi Shahraki, A., & Naseri, A. A. (2017). Simulation of groundwater quality parameters using ANN and ANN+ PSO models (Case study: Ramhormoz Plain). *Pollution*, *3*(2), 191-200.

Songmuang, R., Charusiri, P., Choowong, M., Won-In, K., Takashima, I., & Kosuwan, S. (2007). Detecting active faults using remote-sensing technique: A case study in the Sri Sawat Area, western Thailand. *Sci. Asia*, *33*, 23-33. 10.2306/scienceasia1513-1874.2007.33.023.

Sridharan, M., & Senthil Nathan, D. (2017). Groundwater quality assessment for domestic and agriculture purposes in Puducherry region. *Applied water science*, *7*(7), 4037-4053. 10.1007/s13201-017-0556-y.

Srirattana, S., Piaowan, K., Imthieang, T., Suk-in, J., & Phenrat, T. (2021). Assessment of lead (Pb) leakage from abandoned mine tailing ponds to Klity Creek, Kanchanaburi Province, Thailand. *GeoHealth*, *5*(5), e2020GH000252. https://doi.org/10.1029/2020GH000252.

Srisuk, K., & Nettasana, T. (2016). Climate change and groundwater resources in Thailand. *Journal of* Groundwater Science and Engineering Vol, 4(2).

Strahler, A.N., 1964 (1964). Quantitative geomorphology of basin and channel networks: handbook of applied hydrology.

Suganthi, S., Elango, L., & Subramanian, S. K. (2013). Groundwater potential zonation by Remote Sensing and GIS techniques and its relation to the Groundwater level in the Coastal part of the Arani and Koratalai River Basin, Southern India. *Earth Sciences Research Journal*, *17*(2). 10.15446/esrj.

Sun, J., He, F., Zhang, Z., Shao, H., Pan, Y., Yang, R., Li, W., Li, P., & Zheng, M. (2018). Analysis of saline groundwater infiltration into two loam soils. *Land degradation & development*, *29*(10), 3795-3802.

Tahmassebipoor, N., Rahmati, O., Noormohamadi, F., & Lee, S. (2016). Spatial analysis of groundwater potential using weights-of-evidence and evidential belief function models and remote sensing. *Arabian Journal of Geosciences*, *9*(1), 1-

18. 10.1007/s12517-015-2166-z.

Tamiru, H., & Dinka, M. O. (2021). Application of ANN and HEC-RAS model for flood inundation mapping in lower Baro Akobo River Basin, Ethiopia. *Journal of Hydrology: Regional Studies*, *36*, 100855. https://doi.org/10.1016/j.ejrh.2021.100855.

Tan, Z., Yang, Q., & Zheng, Y. (2020). Machine learning models of groundwater Arsenic spatial distribution in Bangladesh: influence of holocene sediment depositional history. *Environmental Science & Technology*, *54*(15), 9454-9463. https://doi.org/10.1021/acs.est.0c03617.

Tayyebi, A., & Pijanowski, B. C. (2014). Modeling multiple land use changes using ANN, CART and MARS: Comparing tradeoffs in goodness of fit and explanatory power of data mining tools. *International Journal of Applied Earth Observation and Geoinformation*, *28*, 102-116. https://doi.org/10.1016/j.jag.2013.11.008.

Termeh, S. V. R., Khosravi, K., Sartaj, M., Keesstra, S. D., Tsai, F. T. C., Dijksma, R., & Pham, B. T. (2019). Optimization of an adaptive neuro-fuzzy inference system for groundwater potential mapping. *Hydrogeology Journal*, *27*(7), 2511-2534. https://doi.org/10.1007/s10040-019-02017-9.

Thapa, R., Gupta, S., Guin, S., & Kaur, H. (2017). Assessment of groundwater potential zones using multi-influencing factor (MIF) and GIS: a case study from Birbhum district, West Bengal. *Applied Water Science*, *7*(7), 4117-4131. 10.1007/s13201-017-0571-z.

Thapa, R., Gupta, S., Gupta, A., Reddy, D. V., & Kaur, H. (2018). Use of geospatial technology for delineating groundwater potential zones with an emphasis on water-table analysis in Dwarka River basin, Birbhum, India. *Hydrogeology Journal*, *26*(3), 899-922. https://doi.org/10.1007/s10040-017-1683-0.

Tiankao, W., & Chotpantarat, S. (2018). Risk assessment of arsenic from contaminated soils to shallow groundwater in Ong Phra Sub-District, Suphan Buri Province, Thailand. *Journal of Hydrology: Regional Studies*, *19*, 80-96. https://doi.org/10.1016/j.ejrh.2018.08.001.

Tien Bui, D., Shirzadi, A., Chapi, K., Shahabi, H., Pradhan, B., Pham, B. T., Signh, V. P., Chen, W., Khosravi, K., Ahmad, B. B., & Lee, S. (2019). A hybrid computational intelligence approach to groundwater spring potential mapping. *Water*, *11*(10), 2013. https://doi.org/10.3390/w11102013.

Twarakavi, N. K., & Kaluarachchi, J. J. (2005). Aquifer vulnerability assessment to heavy metals using ordinal logistic regression. *Groundwater*, *43*(2), 200-214. https://doi.org/10.1111/j.1745-6584.2005.0001.x.

Twumasi, Y. A., & Merem, E. C. (2006). GIS and remote sensing applications in the assessment of change within a coastal environment in the Niger Delta region of Nigeria. *International journal of environmental research and public health*, *3*(1), 98-106. https://doi.org/10.3390/ijerph2006030011.

Uliana, M. M. (2005). Storage coefficient. *Water Encyclopedia*, *5*, 480-483. https://doi.org/10.1002/047147844X.gw1103.

Uliasz-Misiak, B., Winid, B., Lewandowska-Śmierzchalska, J., & Matuła, R. (2022). Impact of road transport on groundwater quality. *Science of The Total Environment*, *824*, 153804. https://doi.org/10.1016/j.scitotenv.2022.153804.

Velasquez, M., & Hester, P. T. (2013). An analysis of multi-criteria decision making methods. *International journal of operations research*, *10*(2), 56-66.

Wahyudi, A., Bartzke, M., Küster, E., & Bogaert, P. (2013). Maximum entropy estimation of a Benzene contaminated plume using ecotoxicological assays. *Environmental pollution*, *172*, 170-179.

https://doi.org/10.1016/j.envpol.2012.08.018.

Waikar, M. L., & Nilawar, A. P. (2014). Identification of groundwater potential zone using remote sensing and GIS technique. *International Journal of Innovative Research in Science, Engineering and Technology*, *3*(5), 12163-12174.

Wali, S. U., Umar, K. J., Abubakar, S. D., Ifabiyi, I. P., Dankani, I. M., Shera, I. M., & Yauri, S. G. (2019). Hydrochemical characterization of shallow and deep groundwater in Basement Complex areas of southern Kebbi State, Sokoto Basin, Nigeria. *Applied Water Science*, *9*(8), 1-36. https://doi.org/10.1007/s13201-019-1042-5.

Wang, F., Wang, Y., Zhang, K., Hu, M., Weng, Q., & Zhang, H. (2021). Spatial heterogeneity modeling of water quality based on random forest regression and model interpretation. *Environmental Research*, *202*, 111660. https://doi.org/10.1016/j.envres.2021.111660.

Wang, H., Nie, L., Xu, Y., Du, C., Zhang, T., & Wang, Y. (2018). Effects of highway-related pollutant on the groundwater quality of turfy swamps in the Changbai Mountain Area. *International Journal of Environmental Research and Public Health*, *15*(8), 1652. https://doi.org/10.3390/ijerph15081652.

Wang, J. J., Jing, Y. Y., Zhang, C. F., & Zhao, J. H. (2009). Review on multi-criteria decision analysis aid in sustainable energy decision-making. *Renewable and sustainable energy reviews*, *13*(9), 2263-2278. https://doi.org/10.1016/j.rser.2009.06.021.

Wang, Y., Sun, D., Wen, H., Zhang, H., & Zhang, F. (2020). Comparison of random forest model and frequency ratio model for landslide susceptibility mapping (LSM) in Yunyang County (Chongqing, China). *International journal of environmental research and public health*, *17*(12), 4206. https://doi.org/10.3390/ijerph17124206.

White, I., Falkland, T., Metutera, T., Metai, E., Overmars, M., Perez, P., & Dray, A. (2007). Climatic and Human Influences on Groundwater in Low AtollsAll rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. *Vadose Zone Journal*, *6*(3), 581-590. https://doi.org/10.2136/vzj2006.0092.

Wilson, A. E., Sarnelle, O., & Tillmanns, A. R. (2006). Effects of cyanobacterial toxicity and morphology on the population growth of freshwater zooplankton: Meta-analyses of laboratory experiments. *Limnology and Oceanography*, *51*(4), 1915-1924. https://doi.org/10.4319/lo.2006.51.4.1915.

Wisittammasri, W., & Chotpantarat, S. (2016). Isotope evidence of rainfall and groundwater for tracing recharge areas in Kaeng Khoi district, Saraburi province, Thailand. *Applied Environmental Research*, *38*(3), 49-58. https://doi.org/10.35762/AER.2016.38.3.5.

Wisitthammasri, W., Chotpantarat, S., & Thitimakorn, T. (2020). Multivariate statistical analysis of the hydrochemical characteristics of a volcano sedimentary aquifer in Saraburi Province, Thailand. *Journal of Hydrology: Regional Studies*, *32*, 100745. https://doi.org/10.1016/j.ejrh.2020.100745.

Worqlul, A. W., Jeong, J., Dile, Y. T., Osorio, J., Schmitter, P., Gerik, T., Srinivasan, R., & Clark, N. (2017). Assessing potential land suitable for surface irrigation using groundwater in Ethiopia. *Applied Geography*, *85*, 1-13. https://doi.org/10.1016/j.apgeog.2017.05.010.

Wu, C., Fang, C., Wu, X., Zhu, G., & Zhang, Y. (2021). Hydrogeochemical characterization and quality assessment of groundwater using self-organizing maps in the Hangjinqi gasfield area, Ordos Basin, NW China. *Geoscience*

Frontiers, *12*(2), 781-790. https://doi.org/10.1016/j.gsf.2020.09.012.

Wu, C., Wu, X., Qian, C., & Zhu, G. (2018). Hydrogeochemistry and groundwater quality assessment of high fluoride levels in the Yanchi endorheic region, northwest China. *Applied Geochemistry*, *98*, 404-417. https://doi.org/10.1016/j.apgeochem.2018.10.016.

Xie, X., Wang, Y., Ellis, A., Liu, C., Duan, M., & Li, J. (2014). Impact of sedimentary provenance and weathering on arsenic distribution in aquifers of the Datong basin, China: constraints from elemental geochemistry. *Journal of Hydrology*, *519*, 3541-3549. https://doi.org/10.1016/j.jhydrol.2014.10.044.

Xu, Q., Wang, Q., Liu, J., & Liang, H. (2021). Simulation of land-use changes using the partitioned ANN-CA model and considering the influence of land-use change frequency. *ISPRS International Journal of Geo-Information*, *10*(5), 346. https://doi.org/10.3390/ijgi10050346.

Yan, F., Zhang, Q., Ye, S., & Ren, B. (2019). A novel hybrid approach for landslide susceptibility mapping integrating analytical hierarchy process and normalized frequency ratio methods with the cloud model. *Geomorphology*, *327*, 170-187. https://doi.org/10.1016/j.geomorph.2018.10.024.

Yeh, P. J. F., Swenson, S. C., Famiglietti, J. S., & Rodell, M. (2006). Remote sensing of groundwater storage changes in Illinois using the Gravity Recovery and Climate Experiment (GRACE). *Water Resources Research*, *42*(12). https://doi.org/10.1029/2006WR005374.

Zabihi, M., Pourghasemi, H. R., Pourtaghi, Z. S., & Behzadfar, M. (2016). GIS-based multivariate adaptive regression spline and random forest models for groundwater potential mapping in Iran. *Environmental Earth Sciences*, *75*(8), 1-19.

Zandi, J., Ghazvinei, P. T., Hashim, R., Yusof, K. B. W., Ariffin, J., & Motamedi, S. (2016). Mapping of regional potential groundwater springs using logistic regression statistical method. *Water Resources*, *43*(1), 48-57. 10.1134/S0097807816010097.

Zannou, A., & Boulaalam, A. (2021). Relevant node discovery and selection approach for the Internet of Things based on neural networks and ant colony optimization. *Pervasive and Mobile Computing*, *70*, 101311. https://doi.org/10.1016/j.pmcj.2020.101311.

Zehtabiyan-Rezaie, N., Alvandifar, N., Saffaraval, F., Makkiabadi, M., Rahmati, N., & Saffar-Avval, M. (2019). A solar-powered solution for water shortage problem in arid and semi-arid regions in coastal countries. *Sustainable Energy Technologies and Assessments*, *35*, 1-11. https://doi.org/10.1016/j.seta.2019.05.015.

Zhang, X., Liu, L., & Henebry, G. M. (2019). Impacts of land cover and land use change on long-term trend of land surface phenology: a case study in agricultural ecosystems. *Environmental Research Letters*, *14*(4), 044020. 10.1088/1748-9326/ab04d2.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

SUPPLEMENTATION

Supplementary Table 1: Statistic of previous publications relevant to groundwater quality assessment

Publication	K^+	Na^+	Ca^{2+}	Mg^{2+}	Cl^-	SO_4^{2-}	HCO_3^-	NO_3	pH	TDS	TH	EC	F^-	CO_3^{2-}	Fe	Mn	T	ORP	PO_4^{3-}
Konkul et al. (2014)	•	•	•	•	•	•	•		•	•		•		•					
Amiri et al. (2014)	•	•	•	•	•	•	•	•	•	•		•	•						
Sadat-Noori et al. (2014)	•	•	•	•	•	•	•		•	•	•	•							
Boateng et al. (2016)	•	•	•	•	•	•	•	•	•	•	•	•							•
Sridharan and Senthil Nathan (2017)	•	•	•	•	•	•	•	•	•	•	•	•							
Rabeiy (2018)	•	•	•	•	•	•	•		•	•	•	•			•	•			
Li et al. (2018)	•	•	•	•	•	•	•	•	•	•	•	•							
Kawo and Karuppannan (2018)	•	•	•	•	•	•	•		•	•	•	•	•						
Gulgundi and Shetty (2018)	•	•	•	•	•	•	•	•	•	•	•	•	•		•				
Asadi et al. (2019)	•	•	•	•	•	•	•		•	•		•		•					
Wali et al. (2019)	•	•	•	•	•	•		•	•		•	•							
Jha et al. (2020)	•	•	•	•	•	•	•		•	•	•	•	•						
Wu et al. (2021a)	•	•	•	•	•	•	•	•	•			•	•						
Agrawal et al. (2021)	•	•	•	•	•	•	•	•	•	•		•							
Chotpantarat and Thamrongsrisakul (2021)			•	•					•	•	•	•					•	•	
% of parameters	93	93	100	100	93	93	80	53	87	67	53	80	33	13	13	7	7	7	7
This study	•	•	•	•	•	•	•	•	•	•	•	•					•	•	•

Supplementary Table S2: Information on groundwater samples

ID	EC (mS/cm)	TH (mg/l)	TDS (mg/l)	pH	Ca^{2+} (mg/l)	Mg^{2+} (mg/l)	Na$^+$ (mg/l)	K$^+$ (mg/l)	Cl$^-$ (mg/l)	HCO$_3^-$ (mg/l)	SO$_4^{2+}$ (mg/l)	NO$_3^-$ (mg/l)	EWQI normalized	E	N
1	265	216	414	7.25	59.2	16.6	53.8	8.3	21.76	340.46	14.34	4.36	0.15	549107	1615363
2	1559	420	897	7.15	96	43.86	173.57	0.93	256.22	488.03	36.46	1.01	0.3	551760	1569121
3	822	236	373	6.87	78.4	9.82	31.04	2.1	6.99	355.14	3.79	1.27	0.08	550303	1607642
4	1739	300	423	6.9	84	21.97	35.63	4.4	26.42	399.16	9.75	1.37	0.16	545342	1559395
5	1826	570	1024	7.26	62	100.92	183.92	11.31	211.12	795.3	35.91	4.54	0.51	544171	1531022
6	527	380	549	7	126	15.95	53.57	2.58	24.87	496.02	41.76	5.19	0.18	544365	1596099
7	2480	425	602	7.37	104	40.22	66.21	3.07	59.25	572.03	11.49	0.84	0.24	528797	1527935
8	3470	620	931	6.96	106	86.39	125.76	6.55	175.65	798.32	9.68	0.0001	0.44	541421	1535662
9	4400	290	1057	7.18	80	21.97	301.17	0.78	123.96	804.19	100.97	0.13	0.44	551378	1553884
10	627	490	625	7.02	148	29.34	40	4.34	69.17	534.17	33.5	0.26	0.21	548603	1602359
11	1980	425	521	6.85	132	23.25	23.45	3.46	41.19	472.54	11.44	18.57	0.26	539587	1538952
12	998	248	317	7.05	52	28.74	19.31	3	15.2	300.61	16.63	0.49	0.11	540550	1528113
13	918	370	634	7.28	122	15.94	46.67	10.1	43.13	422.64	43.99	23.38	0.31	551141	1592297
14	838	405	522	7.1	100	37.79	26.44	7.19	36.55	488.03	16.24	0.51	0.2	548211	1569775
15	1216	188	341	7.38	50.4	15.13	37.93	2.5	6.22	328.72	2.57	1.1	0.09	553864	1543798
16	2970	470	683	7.01	162	15.99	37.93	3.65	109.97	428.51	42.57	0.73	0.24	541255	1562211

17	1197	345	716	7.06	96	25.63	149.44	0.58	26.04	698.53	36.76	0.5	0.23	548035	1624029
18	883	350	506	7.04	128	7.45	40.92	6.63	67.26	385.6	19.84	7.76	0.19	551153	1610664
19	1060	340	601	6.63	92	26.84	66.21	3.27	90.16	399.16	20.13	21.99	0.26	549700	1581290
20	837	305	546	7.22	92	18.34	71.27	2.61	34.21	472.96	7.62	15.1	0.2	550130	1606060
21	1554	270	380	7.17	80	17.11	17.24	1.82	30.79	280.18	27.34	13.39	0.16	553204	1560336
22	1635	270	444	7.07	92	9.83	46.21	7	37.42	335.63	41.02	0.3	0.17	550460	1593946
23	1924	325	412	7.09	94	21.98	15.63	5.54	42.49	350.22	12.25	14.15	0.21	551108	1579704
24	3.25	178	863	7.55	22.4	29.67	283.47	3.83	9.36	933.92	25.31	0.24	0.29	540024	1532569
25	420	156	290	7.17	40	13.66	27.59	13.49	30.72	204.85	16.26	1.45	0.14	551068	1616697
26	1288	320	822	7.47	72	34.11	173.57	14	89.77	669.18	27.61	3.81	0.36	553736	1583948
27	2240	435	601	7.36	138	22.04	27.82	7.4	41.71	478.63	26.98	19.82	0.31	552544	1554712
28	1369	370	904	7.36	84	38.98	137.94	56.5	156.22	443.19	93.91	55.25	0.92	553658	1576801
29	2110	320	593	7.38	88	24.41	104.83	0.86	40.41	522.43	41.56	11.88	0.26	553206	1551669
30	1092	525	687	6.98	156	32.99	40.92	7.02	47.04	602.5	41.02	13.48	0.31	523506	1580424
31	1180	345	763	7.21	104	20.78	130.12	5.05	83.59	463.93	74.99	60.98	0.51	551808	1572024
32	1308	232	362	7.45	84	5.45	37.24	2.68	27.68	326.87	2	0.87	0.09	548427	1556097
33	1447	320	896	7.13	96	19.56	224.84	2.12	85.49	751.36	57.99	1.33	0.31	552306	1602080
34	810	430	470	7.14	140	19.61	0.0001	2.89	15.16	451.88	9.04	31.6	0.26	542376	1563997
35	666	148	237	7.13	49.6	5.89	7.13	1.35	5.85	169.27	13.31	1.02	0.02	539781	1620890
36	237	265	656	7.22	62	26.81	122.54	2.08	44.44	478.63	64.05	6.53	0.22	543284	1605234

37	2130	40	130	5.98	9.6	3.9	2.76	3.42	9.36	23.35	0.77	19.57	0.1	552692	1596986
38	1155	410	650	7.2	68	58.4	83.91	11.9	44.71	611.54	45.18	8.32	0.36	553204	1624982
39	1605	360	470	6.87	76	41.4	28.05	1.97	16.32	460.8	8.04	11.56	0.21	548391	1540284
40	414	445	604	7.01	108	42.66	68.05	4.7	27.2	648.64	14.15	0.41	0.22	545234	1610541
41	2060	470	525	7.14	108	48.73	7.36	1.06	10.88	551.78	5.34	12.11	0.24	537571	1542106
42	1648	370	436	7.24	72	46.26	10.12	1.66	6.99	454.93	5.59	0.65	0.15	540314	1528808
43	1387	560	919	6.87	200	14.82	71.5	7.8	113.08	498.95	37.36	137.23	0.84	547718	1589105
44	904	360	560	7.1	110	20.79	60.46	3.56	47.8	457.86	32.94	14.6	0.23	543399	1594038
45	3410	214	412	7.41	56	18.05	72.19	0.83	9.72	416.77	7.44	2.56	0.17	542213	1541697
46	1394	308	348	6.98	96	16.64	7.59	2.54	9.75	367.73	3.94	5.37	0.12	525453	1576832
47	2330	520	642	7.08	190	11.17	14.02	2.25	26.81	469.6	128.8	4.47	0.28	516729	1536127
48	342	455	594	6.92	124	35.39	34.03	6.78	4.27	604.61	23.99	0.68	0.21	546547	1613505
49	1875	270	458	7.12	88	12.26	46.67	1.91	17.1	369.81	35.69	2.77	0.15	541855	1540247
50	1439	204	309	7.99	40	25.32	31.73	4.54	63.54	169.27	40.47	0.41	0.13	538545	1520826
51	1149	535	587	6.9	142	43.91	23.22	1	37.69	619.29	12.43	0.38	0.2	550699	1619219
52	848	420	453	7.03	108	36.58	0.23	3.05	16.32	463.73	12.39	0.16	0.15	542402	1601468
53	3580	495	939	7.05	68	79.06	191.28	6.08	91.32	939.2	12.39	0.49	0.45	524911	1539359
54	2600	485	605	7.13	104	54.8	40.23	2.04	45.22	513.66	76.79	1.6	0.29	534611	1525667
55	896	440	542	7.52	56	72.97	30.58	6.45	30.4	577.86	0.26	0.47	0.24	538699	1524369
56	1820	470	467	7.12	126	37.82	1.15	0.94	6.99	557.65	1.02	0.26	0.17	540302	1558451

57	1859	370	518	7.1	104	26.86	32.65	1.92	23.32	466.67	17.11	3.6		0.18	531550	1563615
58	1567	365	449	7.17	120	15.94	23.45	0.75	39.64	396.23	25.72	3.01		0.14	536444	1557577
59	1960	324	387	6.94	115.2	8.89	17.7	0.31	21.83	379.41	3.94	16.62		0.17	543982	1554577
60	1847	380	462	7.11	98	32.92	28.28	0.19	24.48	481.34	4.08	5.52		0.17	549071	1545154
61	2050	485	553	6.95	108	52.38	23.45	10	25.65	592.87	14.62	4.2		0.3	547798	1613103
62	1019	425	668	7.03	94	46.28	78.4	13.26	40.82	572.38	69.76	2.74		0.34	545329	1621849
63	1051	395	550	6.93	136	13.53	37.24	3.45	13.21	490.15	54.28	0.08		0.18	542738	1628646
64	1012	535	773	6.96	132	49.98	58.62	2.68	25.65	581.13	140.51	0.83		0.32	530170	1531235
65	1826	560	1371	6.87	104	73.03	252.89	1.53	28.37	584.07	585.11	0.38		0.76	543538	1598507
66	803	220	630	7.39	62.4	15.63	143.23	2.35	47.41	463.73	58.39	0.39		0.19	546888	1624607
67	2.41	370	708	7.27	80	41.41	115.41	3.07	40.93	461.12	167.17	0.19		0.29	525418	1569819
68	1733	380	431	6.93	122	18.37	5.75	2.87	16.76	423.18	29.44	0.11		0.15	540861	1607541
69	2.21	390	527	7.37	92	38.99	48.28	5.45	77.96	434.86	24.7	0.19		0.18	528560	1527930
70	1829	325	352	7.01	92	23.2	2.53	0.79	13.25	367.73	12.32	0.84		0.11	538212	1563868
71	1958	410	443	7.16	84	48.7	12.41	1.32	28.46	472.8	6.5	0.33		0.17	547684	1525711
72	1898	325	344	6.91	112	11.07	4.83	0.39	8.19	370.65	4.2	6.64		0.12	542552	1542306
73	1813	355	458	7.09	126	9.87	21.15	3.07	19.1	361.89	73.23	0.05		0.17	516697	1535858
74	1555	375	361	7.28	70	48.69	0.23	0.37	7.02	417.35	8.05	1.9		0.14	536940	1542710
75	853	440	525	7.08	136	24.47	12.64	6.57	23.33	494.05	26.93	14.62		0.25	538741	1568355
76	1822	330	411	6.9	115.2	10.35	20.23	1.45	12.47	396.92	5.74	22.67		0.21	534690	1572107

77	2940	545	780	6.83	102	70.59	76.79	2.85	90.54	657.44	38.91	1.38	0.34	534644	1524115
78	2840	480	733	7	126	40.25	97.02	2.7	94.43	672.12	6.28	8.43	0.32	530130	1540675
79	962	248	472	7.43	57.6	25.34	88.28	2.23	35.75	463.73	6.9	0.93	0.15	551965	1625626
80	614	525	695	6.96	136	45.12	54.26	0.61	85.1	569.39	58.41	0.38	0.23	542167	1620045
81	2920	285	951	7.13	56	35.3	232.2	3.9	45.85	478.41	293.74	6.74	0.52	544383	1528042
82	3380	450	736	7.22	98	49.93	69.43	15.8	24.87	545.91	140.84	2.71	0.46	537905	1548632
83	686	250	555	7.34	68	19.52	100.24	5.45	56.35	425.58	34.99	0.18	0.18	541856	1608755
84	2240	450	581	7.11	96	51.15	55.41	0.25	19.43	613.42	29.19	2.68	0.24	534149	1537178
85	1982	358	451	7.11	91.2	31.7	24.83	1.9	39.37	414.43	8.95	7.46	0.19	550813	1562141
86	1624	410	448	7.14	140	14.75	7.36	1.47	7.02	484.47	8.15	8.24	0.17	533433	1573942
87	864	395	525	6.93	120	23.23	39.08	2.39	14	524.18	16.24	23.96	0.25	537421	1570901
88	2140	480	582	7.04	146	28.12	7.59	1.24	29.15	507.76	17.04	22.75	0.27	527716	1568553
89	698	380	397	7.27	122	18.37	0.0001	0.38	5.83	451.88	2.38	1.33	0.09	536548	1573035
90	6030	1170	1720	7.06	316	92.72	133.57	2.17	293.13	557.43	539.23	16.94	1	525334	1569719
91	1832	385	483	6.93	150	2.61	33.34	0.91	5.83	522.43	5.5	11.61	0.18	526864	1560455
92	1934	390	390	7.09	88	41.42	0.23	0.53	6.61	422.64	10.42	20.62	0.23	541494	1552381
93	3720	575	936	7.08	158	43.93	84.14	5.48	222.28	501.89	29.65	3.44	0.37	548962	1532570
94	2320	455	676	6.88	176	3.86	73.57	0.17	31.09	610.48	43.82	21.17	0.3	528921	1558273
95	846	345	410	7.14	104	20.78	14.94	3.96	17.88	403.68	17.89	5.55	0.15	540178	1566086
96	2650	500	508	6.99	114	52.38	10.35	0.48	12.05	598.74	2.37	8.82	0.25	545713	1553631

97	542	106	132	7.11	24	11.21	4.14	1.78	3.11	126.21	1.1	3.66	0.02	539148	1561451
98	5490	430	563	7.06	120	31.74	14.02	1.21	43.13	413.84	4.82	59.35	0.48	547862	1543131
99	897	212	312	7.35	74.4	6.41	12.64	1.81	7.02	268.5	6.3	0.39	0.05	539617	1620895
100	3920	550	1117	7.13	204	9.97	154.95	5.11	129.79	425.58	327.97	10.85	0.58	536132	1546603
101	4210	255	948	7.45	28	44.99	258.18	0.48	45.08	851.15	21.86	18.6	0.44	532173	1560475
102	567	302	368	6.64	84	22.46	15.17	0.78	5.44	384.49	2.27	0.33	0.08	542734	1576985
103	804	268	471	7.24	68	23.9	72.42	9.5	32.66	439.83	16.45	0.41	0.19	551030	1617565
104	2390	475	684	7.1	102	53.58	49.43	3.37	32.64	554.72	22.25	69.01	0.54	535110	1540194
105	2530	525	663	7.03	84	76.65	46.21	11.3	80.83	498.95	98.98	0.47	0.4	534079	1528813
106	1020	490	641	7.1	150	28.13	31.27	2.14	54.4	525.37	11.32	41.81	0.35	545671	1586285
107	2630	565	861	6.92	136	54.84	96.79	2.35	33.81	663.31	172.93	0.52	0.41	544536	1625654
108	3790	465	887	7.03	110	46.3	140.24	0.85	223.83	542.98	15.88	7.62	0.36	548820	1543904
109	334	465	593	6.96	120	40.24	32.65	6.7	6.99	619.29	16.98	0.86	0.21	545696	1614880
110	3980	640	837	6.87	150	64.58	60.46	1.25	120.08	557.65	124.92	12.22	0.45	536412	1550781
111	1722	380	440	7.24	107.2	27.35	14.02	1.37	16.37	431.94	28.25	13.75	0.21	552197	1556721
112	833	190	303	7.52	67.2	5.43	11.72	0.53	4.68	245.15	3.28	0.29	0.03	533800	1621237
113	1027	510	776	7.32	138	40.26	72.42	7.95	22.94	578.4	168	0.38	0.37	535608	1629572
114	451	70	191	6.45	6.4	13.13	14.02	4.69	13.6	99.79	0.58	0.57	0.02	552558	1533531
115	1480	310	384	6.95	96	17.13	17.93	2.98	25.65	328.72	40.65	0.13	0.14	527229	1524613
116	2080	380	412	7.1	86.4	39.96	2.76	1.4	22.61	402.75	6.28	17.68	0.22	553132	1559523

117	919	360	597	7.6	110	20	71	3.8	59	433	52	0.52	0.19	589004	1534361
118	789	380	519	7.4	120	19	26	2.8	53	398	17	17	0.2	583224	1530403
119	488	240	317	7.7	73	13	17	1.6	16	266	23	3.4	0.08	581364	1537629
120	605	230	393	7.5	74	11	42	1.5	54	260	28	1.6	0.09	581144	1541135
121	724	350	471	7.5	110	20	20	2.2	30	342	51	21	0.22	585468	1543304
122	491	250	319	8	75	15	17	2.4	12	315	6	0.34	0.07	587503	1547110
123	756	320	491	7.7	89	23	55	3.1	31	391	41	14	0.21	581939	1549614
124	461	120	300	8.6	9.3	24	72	1.1	10	245	22	0.26	0.07	590191	1554297
125	449	230	292	7.9	72	11	12	2.1	9.2	284	7	1.3	0.06	583186	1555574
126	618	89	402	9	13	14	100	10	81	183	5	0.45	0.13	566859	1550993
127	937	380	609	7.7	100	28	74	1.4	40	466	59	17	0.26	567820	1546039
128	717	390	466	7.8	130	14	6	7.4	25	431	5	18	0.22	561668	1548337
129	535	240	348	8.1	66	19	36	2.8	15	302	31	12	0.15	574952	1546758
130	1650	560	1070	7.5	170	31	150	2.7	250	600	37	19	0.4	572001	1540385
131	855	360	556	7.8	110	22	42	7.5	68	402	37	12	0.24	573835	1534797
132	1000	480	650	7.6	140	32	29	6.1	79	532	17	4.4	0.23	567841	1527681
133	925	490	601	7.5	140	32	16	1.3	46	533	4	28	0.27	562568	1527616
134	1220	510	793	8	180	15	59	3.1	140	468	34	48	0.41	557382	1527560
135	902	310	586	7.5	120	0.2	21	1.5	82	283	14	31	0.21	553434	1525529
136	1140	440	741	7.8	130	26	45	1.8	59	378	88	18	0.28	566677	1534978

137	1210	420	786	7.7	130	21	95	2.4	54	398	210	14	0.37	566073	1539606
138	1600	580	1040	7.4	160	47	110	2.3	210	405	67	81	0.65	555307	1535683
139	737	280	479	7.8	78	21	43	1.6	13	395	29	2.3	0.13	556781	1548392
140	722	250	469	8.1	81	11	68	3.1	18	449	6	0.11	0.12	561589	1551541
141	268	130	174	7.9	42	5.4	4.3	1.9	8	157	4	1.2	0.01	559452	1553365
142	1260	340	819	7.5	93	25	140	13	110	460	57	49	0.52	576400	1554652
143	1970	490	1280	7.6	110	49	250	6.6	330	555	57	9.1	0.47	566822	1564739
144	767	290	499	7.7	64	32	64	2.4	50	383	21	2.7	0.16	559489	1565116
145	794	310	516	7.4	95	18	31	5.2	42	371	17	3.2	0.15	557906	1559390
146	1180	340	767	7.8	60	46	150	2.4	13	679	58	0.21	0.28	557701	1581836
147	1000	410	650	7.2	86	47	72	4.6	38	578	41	0.65	0.24	558562	1588802
148	2390	1200	1550	7.2	240	150	65	7.4	320	457	580	0.4	0.94	563190	1583746
149	181	47	118	7.3	16	1.6	20	3.4	7.2	98	3	2.3	0	568539	1582862
150	1050	360	682	7.3	100	25	81	11	31	501	76	1	0.28	574442	1579374
151	1010	550	656	7.4	120	64	30	8	22	630	36	1.9	0.3	574034	1576288
152	1490	280	968	7.8	42	42	280	3	120	754	8	14	0.38	577300	1566592
153	1280	350	832	7.3	110	22	110	24	33	590	84	0.51	0.41	572143	1567391
154	1080	170	702	8.2	35	20	210	4.2	75	599	3	0.9	0.22	579101	1584542
155	1320	490	858	7.6	83	70	100	8.3	120	631	35	17	0.41	569237	1589226
156	581	160	378	7.9	31	21	75	3.6	20	351	2	0.53	0.1	564981	1594558

157	1330	460	864	7.8	85	59	150	6.3	100	771	10	0.35	0.33	572332	1602620
158	2150	970	1400	7	270	71	66	6.1	410	479	68	44	0.65	566356	1606458
159	276	86	179	6.5	28	3.8	16	3.5	19	67	2	53	0.22	559045	1605704
160	752	370	489	7.4	120	20	6	6.4	14	423	6	40	0.31	560648	1608290
161	721	280	469	7.4	84	16	37	3.6	32	337	29	0.44	0.12	557907	1612649
162	1190	560	774	6.9	180	27	41	16	41	745	4	0.62	0.32	560587	1617199
163	1190	260	774	7.4	68	23	160	7.6	11	795	18	0.46	0.27	567342	1613512
164	1790	310	1160	7.5	47	47	340	15	32	1200	2	0.52	0.5	576175	1614023
165	414	120	269	7.1	40	4.6	32	22	30	196	0.67	0.71	0.17	576706	1606723
166	742	340	482	7.5	100	19	18	1.7	18	374	37	10	0.16	583115	1600592
167	840	200	546	8	37	26	110	2.2	28	494	3	0.49	0.15	580890	1588183
168	323	120	210	6.9	28	12	8.2	8.7	40	88	1	17	0.12	572867	1632875
169	772	390	502	7.7	110	32	20	4.9	35	453	4	0.53	0.16	567963	1634355
170	1190	400	774	7.8	58	63	110	17	81	658	20	5.9	0.4	578980	1626483
171	791	230	514	8.1	60	20	70	5.3	25	417	12	0.18	0.14	578818	1626399
172	1100	160	715	7.9	27	23	200	13	57	629	2	0.2	0.28	574775	1622343
173	680	240	442	7.3	38	36	47	11	58	339	4	0.72	0.18	572897	1622568
174	268	76	174	7.3	23	4.4	19	15	19	105	18	2.8	0.11	586178	1618486
175	664	170	432	7.7	52	11	83	9.7	19	404	11	0.32	0.16	584223	1613810
176	759	360	493	7.4	91	31	20	5	8.4	445	16	0.49	0.15	586062	1604964

177	1110	440	722	7.4	130	31	72	1.5	110	448	42	6.8	0.24	586414	1593597
178	269	130	175	7.8	39	7	8.2	1.5	14	148	4	1.8	0.01	582502	1571556
179	910	360	592	7.3	100	28	45	2.5	110	345	14	6.4	0.18	588381	1577241
180	115	41	75	5.6	9.6	4.2	11	2.5	22	24	2	6.6	0	582855	1577733

Structure of algorithm

Algorithm 1. Random forest for predicting GW yield

1: # Install package "randomForest" and "Java environment"

2: # Import databases

3: # Set training and testing dataset

3: # Construct Random Forest model

4: RF = randomForest('environmental factors', 'train data', 'ntree', 'cross-validation')

5: # Get the map

6: plot(RF)

Algorithm 2. Maxent approach

1: # Install package "SDM" and "Java environment"

2: # Import databases

3: # Construct maxent model

4: MX = max('environmental factors', 'train data')

5: # Get the contribution variable

6: plot(MX)

7: # Get the predicted value

8: response(MX)

Algorithm 3. Random forest for deciphering GW quality

1: # Install package "randomForest" and "plyr"

2: # Import databases

3: # Set Cross-validation

4: # Set Bootstrap

5: # Construct RF-CV and RF-B models

6: RF_CV = train('influencing factors', 'dataset', 'RF', 'trControl = method'CV', 'number =5")

7: RF_B = train('influencing factors', 'dataset', 'RF', 'trControl = method'boot',

'number =100")

8: # Determine the best model

9: which.min(plot(RF_CV$finalModel,fallen.leaves = F))

10: which.min(plot(RF_B$finalModel,fallen.leaves = F))

11: Get the maps

12: plot(RF_CV)

13: plot(RF_B)

14: # Determine the important value

15: varImp(RF_CV)

16: varImp(RF_B)

Algorithm 4. Artificial Neural Networks for deciphering GW quality

---

1: # Install package "neuralnet"

2: # Import databases

3: # Set Cross-validation

4: # Set Bootstrap

5: # Construct ANN-CV and ANN-B models

6: ANN_CV = train('influencing factors', 'dataset', 'nnet', 'trControl = method'CV',
'number =5")

7: ANN_B = train('influencing factors', 'dataset', 'nnet', 'trControl = method'boot',
'number =100")

8: Get the maps

9: plot(ANN_CV)

10: plot(ANN_B)

11: # Determine the important value

12: varImp(ANN_CV)

13: varImp(ANN_B)

VITA

NAME   NGUYEN NGOC THANH

DATE OF BIRTH   27 February 1990

PLACE OF BIRTH   Quang Binh - Vietnam

INSTITUTIONS ATTENDED   Interdisciplinary Program in Environmental Science.
Graduate School, Chulalongkorn University

HOME ADDRESS   53 Tran Nguyen Dan, Hue city, Vietnam

PUBLICATION   1) Thanh, N. N., Thunyawatcharakul, P., Ngu, N. H., & Chotpantarat, S. (2022). Global review of groundwater potential models in the last decade: parameters, model techniques, and validation. Journal of Hydrology, 128501.

2) Thanh, N. N., Chotpantarat, S., Trung, N. H., & Ngu, N. H. (2022). Mapping groundwater potential zones in Kanchanaburi Province, Thailand by integrating of analytic hierarchy process, frequency ratio, and random forest. Ecological Indicators, 145, 109591. [Q1, IF 6.2]