

Research Article

Automatic Sinkhole Classification by Employing Field Data: Application to Sinkholes in South Korea

^aWoohyun James Chun, ^bChang Joon Chung, ^cHyungjun Chin, ^dSeongbin Cho, ^eDong Yeon Hu, ^fMadeline Jung and ^gBob Nam

^aChoate Rosemary Hall, USA, ^bSeoul Foreign School, South Korea, ^cSeoul International School, South Korea, ^dAvon Old Farms School, USA, ^{e&f}Cornerstone Collegiate Academy of Seoul, South Korea, ^gSeoul Innovations Research Institute, South Korea

*Corresponding Author Email: bobyep25@gmail.com

Received: September 16, 2025

Accepted: October 07, 2025

Published: October 13, 2025

Abstract

Many complicated factors can play a simultaneous role in sinkhole attacks. The surface and groundwater level can be altered due to human involvement such as infrastructure settings or due to the environmental changes such as global warming or precipitation, along with the complex hydro-chemical properties implicitly governing the changes in ground surface. As for these causes of sinkhole attacks, government of South Korea has adopted causality based six classifications, measuring and deploying all countrywide sinkhole event records for eight years. Moreover, Geo Big Data Platform of Korean government enables the recognition of the underlying geographical characteristics below the sinkhole attack regions. By extending and fusing these data sets into 13 features related with ground subsidence, we applied Random Forest (RF) Classifier for automatic sinkhole classification, obtaining 94.7% accuracy. Especially, machine learning model performance is highly enhanced when geological characteristics of sinkhole attack regions are combined. With this investigation, it can be suggested that field data, independently of the generally utilized thermal images from drones or from GPR, or LiDAR data, can also be comparably useful in identifying and classifying sinkholes, especially for the case of sinkholes in metropolitan areas like in Korea. Merging these field data into the potential data sets from various wireless sensors like 3d GPS to make it more plausible to forecast sinkhole attacks in advance is our subsequent research agenda.

Keywords: Sinkhole, Machine Learning, Random Forest Classifier, Geological Characteristics, Field Data for Sinkholes.

Introduction

Sinkholes, a commonly used term to express collapse dolines, are natural phenomena with sizes and depths ranging from centimeters to meters, occurring in karstic regions containing carbonate and evaporitic rocks that cover approximately 20% of the Earth's surface [15]. Sinkhole formations occur under the control of many topographic, geological-tectonic, environmental anthropogenic, hydrogeological, and climatic (meteorological) factors [22]. These determining factors can be exemplified as the surface and groundwater level that enable dissolution with soluble carbonate and evaporitic karst rocks, the flow direction and hydro-chemical properties of water, precipitation, evaporation, stratification that facilitates the movement of water in rocks, porosity and permeability, and cracks and fractures [15, 16]. However, due to the influence of many distinct factors, it is challenging to model sinkhole formations in temporal and spatial dimensions [17, 18, 19]. For this reason, the modeling of sinkhole susceptibility maps is a complex process, as in other types of natural phenomena, and performing these operations with classical methods can be a very time-consuming and puzzling task for the decision-makers [22]. Deterministic models attempt to analyze sinkhole formation based on hydrogeological and geotechnical data within the framework of physical and mechanical principles [22]. However, practical application of these methods is limited due to the high demand for detailed field data and their restricted scalability over large areas [20, 21].

Sinkholes in Korea, however, reveal a clear tendency to be clustered in the metropolitan areas, enabling the relatively clear-cut estimation of the causes, since the infrastructure settings are more directly involved with the sinkhole formations rather than the underlying hydrogeological and geotechnical processes. The fact that

about 53% of sinkholes during recent three years are caused by water supply or drain-pipe damages represents sinkholes in Korea are relatively easier to identify, classify and manage, since the drastic underground movement related with water flow can dominate slow and normal geotechnical procedures. Moreover, cumulated sinkhole tracking data sets for eight years can shed a light, even in case where the data sets are not complete and through. With these prospects in mind, our study employs publicly available field data sets about sinkholes from data base in Underground Safety Information Network (<https://www.jis.go.kr/>) and geological data sets below the sinkhole attack regions in Geo Big Data Platform (<https://data.kigam.re.kr/>) of Korean government to develop an automatic sinkhole classification model, facilitating complicated on-field classification procedures which normally take several months. Fusing these field data sets into the potential data sets from various sensors is our next step to further our search for an in advance warning system of sinkhole attacks.

Literature Review

Deep learning-based sinkhole detection has emerged as a promising tool to address the challenges in accurately identifying and predicting sinkholes in various environments, such as urban, natural, and infrastructure settings. Many literatures reveal a transition from classical machine learning methods to deep multimodal architectures that merge high-resolution imaging and spatiotemporal data for improved accuracy and robustness [1, 2, 3, 4].

Early Machine Learning Innovations

Initial research focused on classical machine learning techniques such as support vector machines, decision trees, and ensemble models, primarily leveraging simulated or limited real-world datasets for sinkhole detection in Wireless Sensor Networks (WSNs) [1-14]. These approaches aimed for high precision in anomaly detection, with innovations around feature selection and data optimization for energy efficiency and false positive reduction [1].

Advancement to Deep Learning Architectures

Recent studies shifted towards deep learning, utilizing convolutional neural networks (CNNs), multimodal fusion techniques, and weakly supervised learning [2, 3]. The introduction of frameworks like SinkholeNet enabled the classification and localization of sinkholes in high-resolution RGB-slope images using fused multimodal inputs. These deep learning models often excel traditional methods, benefiting from extensive datasets that incorporate pixel-level labels and topographical features, while offering superior spatial understanding and adaptability [2, 3].

Techniques such as Frequency Ratio (FR) [24, 25, 26], Logistic Regression (LR) [21, 27, 28], Linear Discriminant Analysis (LDA) [29], and Weight of Evidence (WoE) [30] have been widely employed in generating sinkhole susceptibility maps.

Application to Remote Sensing and Infrastructure

Modern research encompasses remote sensing data such as LiDAR, InSAR time series, and multispectral images to detect sinkholes over large areas and infrastructure settings, including railways and urban landscapes [4, 5, 6]. Advanced deep learning algorithms are employed to model sinkhole morphology, learn spatiotemporal event patterns, and automate the mapping and assessment tasks. These approaches facilitate earlier and more reliable detection in complex geophysical environments, outperforming previous classical approaches in recall and localization accuracy [4, 5, 6].

Dataset Innovations and Benchmarking

The lack of extensive annotated datasets for sinkhole detection prompted the development of novel open-access datasets, such as RGB-slope patches and synthetic features from simulation environments [2, 3]. These data sets enable benchmarking and comparison of various deep learning algorithms-segmentation models, weakly supervised classifiers, and multimodal fusion techniques-demonstrating significant improvements in overall hazard detection and susceptibility analysis [2, 3, 7].

In summary, literatures indicate that deep learning methodologies provide enhanced performance, flexibility, and scalability for sinkhole detection across environmental and infrastructure domains. Recent advances focus on exploiting multimodal remote sensing and developing comprehensive public datasets, thus supporting more accurate classification, localization, and susceptibility analysis of sinkholes [1, 2, 3, 4]. Table 1 summarizes machine learning (ML) or deep learning (DL) studies to detect sinkholes, a simplified version of Table 1 of [14].

Table 1. Simplified summary of ML or DL studies to detect sinkholes [14].

Authors	Technique used	Data source	Performance metrics
Lee et al. [8]	3D-convolutional neural network (CNN)	Thermal images from drones, resolution: 640×480 pixels.	Precision: 87.9% Recall: 88.1%
Zhu and Pierskalla [9]	Random forest classifier	LiDAR data, average point spacing: 1 m, DEM cell size: 1.5 m.	Precision: 84.71% Recall: 65.17%
Kang et al. [10]	Modified CNN architecture based on AlexNet (Krizhevsky et al. [31])	Ground penetrating radar (GPR), original resolution: 50×50 pixels (B-scan), 50×13 pixels (C-scan), enhanced to 200×200 pixels.	(Original resolution) Precision: 88.26%, Recall: 72.36%, (Enhanced) Precision: 100% Recall: 100%,
Mihevc and Mihevc [11]	U-Net	LiDAR, DEM cell size: 1m	Intersection over union (IoU): 60.4% Dice coefficient: 72.36%
Nefeslioglu et al. [12]	Artificial neural network (ANN)	Satellite optical imagery and InSAR DEMs spatial resolution: 10 m.	Root mean square error (RMSE): 45.1%
Rafique et al. [13]	U-Net	LiDAR DEMs, aerial imagery resolution: 1.524 m per pixel.	IoU: 45.38% Precision: 66.29%

Data

Underground Safety Information Network (<https://www.jis.go.kr/>) of Korean government has recorded all 1554 sinkhole attacks from January 1, 2018 to September 30, 2025, allowing public access and search. For each sinkhole attack in South Korea, date, address, width, extended width, depth, estimated cause, estimated detailed cause, number of casualties, number of damaged cars, restoration status, restoration methods and future restoration plans are recorded. According to this data base, yearly number of sinkhole accidents reveals slightly decreasing pattern, although the pattern is somewhat ambiguous for recent five years (Figure 1).

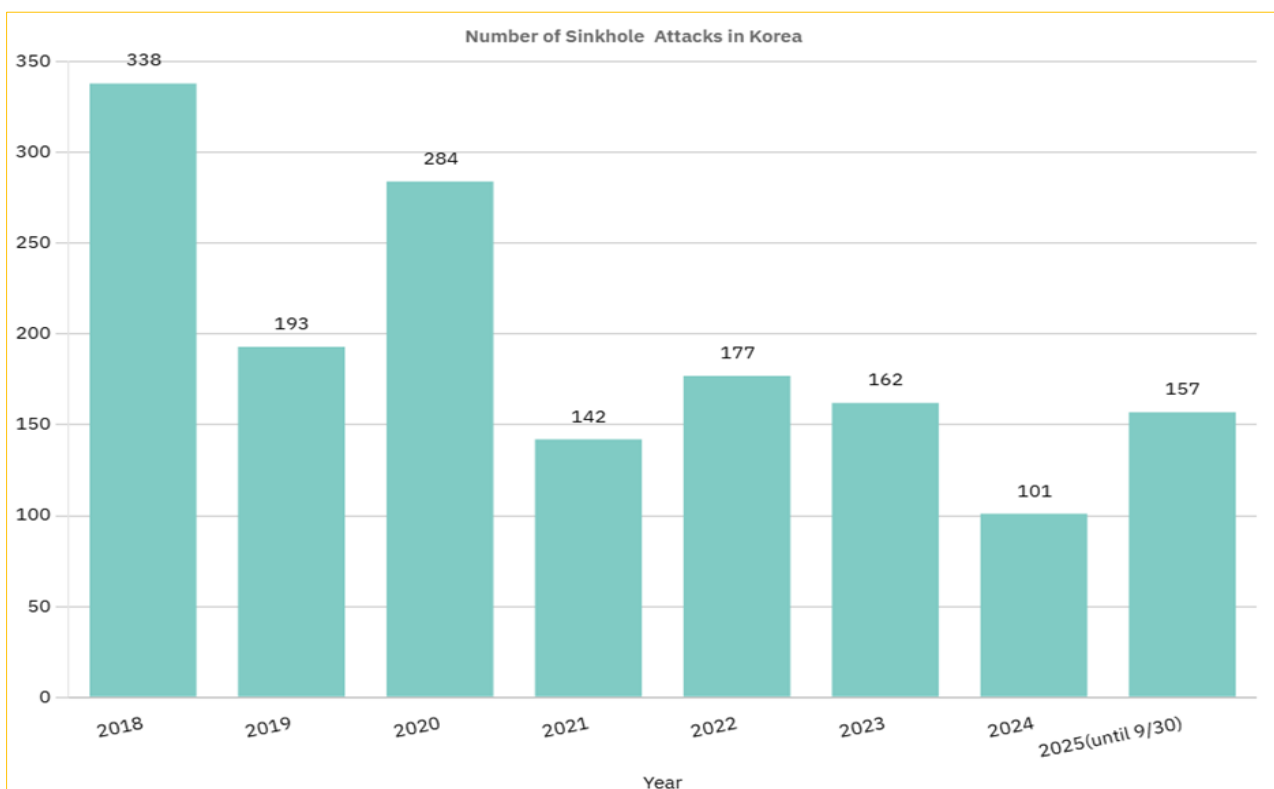
**Figure 1.** Number of sinkhole attacks in Korea.

Table 2 denotes six sinkhole classes with the notation in this study and estimated causes applied in the Underground Safety Information Network site.

Table 2. Sinkhole types.

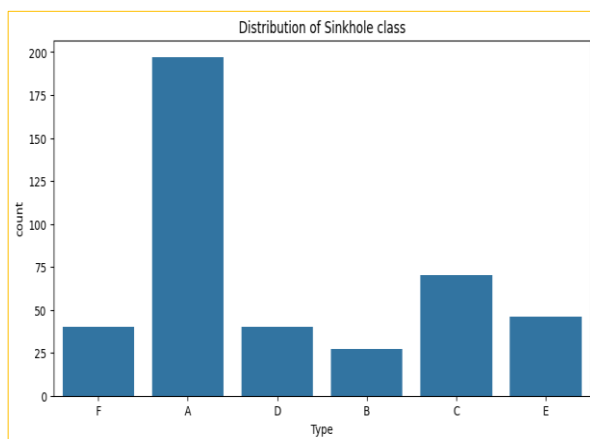
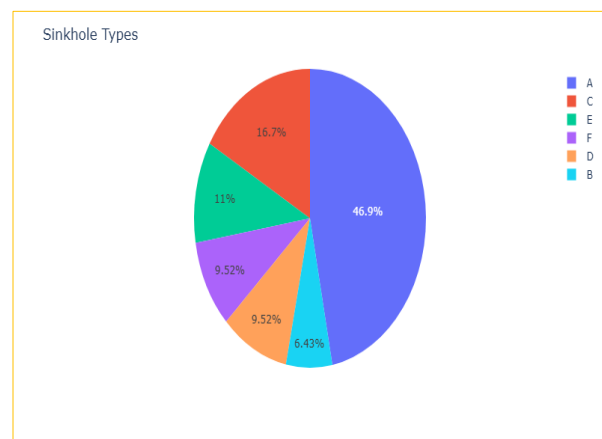
Notation of sinkhole type	Evaluation of the causes of sinkhole events
A (numeric value assigned: 0)	Supply-pipe damage
B (1)	Drain-pipe damage
C (2)	Poor soil compaction
D (3)	Poor excavation
E (4)	Underground facilities damage
F (5)	Complex reasons or indeterminable

Among these 1554 sinkhole disasters, 440 recent cases from January 1, 2022 to December 31, 2024 are included in our study (420 cases for trained and 20 for test sets). Data values of year 2025 are not included since classification has not been completed during the data collection stage. Summary statistics for these 420 observations are given in Table 3. Graphical representations of six sinkhole types are given in Figure 2 and Figure 3.

Table 3. Summary statistics of variables.

	Latitude	Longitude	Width(m)	Width.extension	Depth	Death	Injured	Car Damage	Altitude(meter)	Month
count	420.000000	420.000000	420.000000	420.000000	420.000000	420.000000	420.000000	420.000000	420.000000	420.000000
mean	36.441425	127.526108	2.292095	3.381548	1.533000	0.004762	0.088095	0.109524	52.866595	6.395238
std	1.095194	0.827883	9.859848	6.457423	2.276039	0.068924	0.482981	0.471027	72.702686	2.512209
min	33.393220	125.817800	0.030000	0.000000	0.020000	0.000000	0.000000	0.000000	-8.000000	1.000000
25%	35.227325	126.921225	1.000000	1.000000	0.787500	0.000000	0.000000	0.000000	14.750000	5.000000
50%	36.352875	127.128300	1.500000	1.500000	1.315000	0.000000	0.000000	0.000000	36.000000	7.000000
75%	37.489653	128.221325	2.000000	3.000000	1.900000	0.000000	0.000000	0.000000	60.000000	8.000000
max	38.275590	129.343900	200.000000	55.000000	40.000000	1.000000	8.000000	7.000000	815.000000	12.000000

According to below Figure 3, approximately 53 percent of sinkholes in South Korea for three years are caused by supply water-pipe damages (Type A) or drain water pipe damages (Type B). Except for supply water pipe related class, other types have roughly similar frequencies. Since most sinkhole cases are noticed in metropolitan areas, such as Seoul, Gyeonggi, Pusan, and Gwangju, features related with infrastructure settings might dominate those related with environmental characteristics, the latter of which might be committed to the class of complex and unknown causes (Type F) when field data are registered, since not a geological analysis but a constructional restoration has been the main purpose of sinkhole event record.

**Figure 2.** Bar chart of sinkhole type.**Figure 3.** Pie chart of sinkhole type.

From Figure 4 and Figure 5, about 43% of sinkholes occur within summer seasons from June to August, implying that rainfalls have a major role in the presence of Type A sinkholes. Frequencies of sinkholes tend to increase from February to August, decrease until October, and then remain roughly constant or increase during winter seasons. When the histogram of month is segmented, a clear pattern comes up that the

proportion of sinkholes caused by water supply pipe damage increases from March to August, moving together with the increase in summer rainfalls. The addresses of sinkhole observations are transformed into latitude, longitude and altitude to make it possible to track the geological implications.

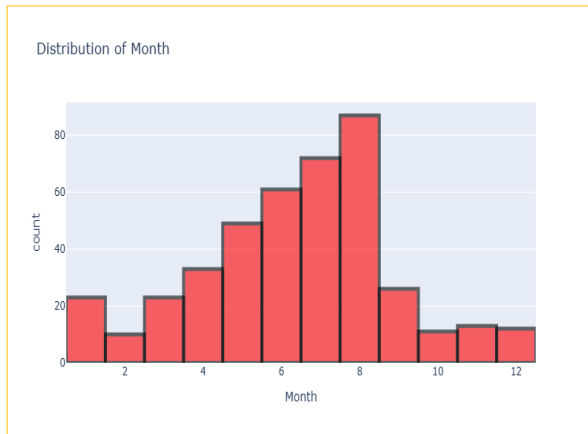


Figure 4. Histogram of month.

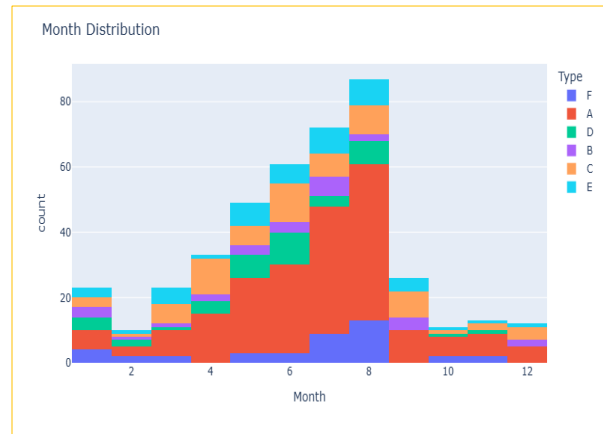


Figure 5. Segmented histogram of month.

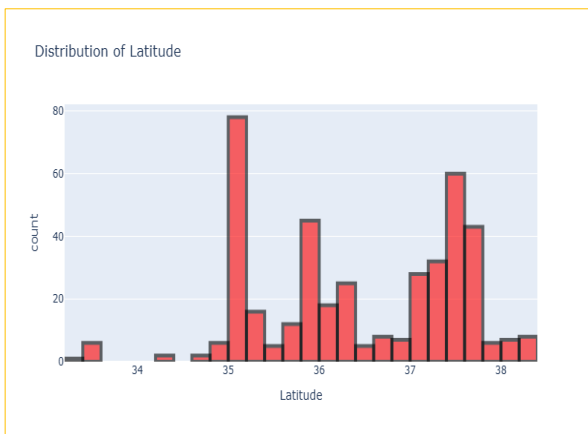


Figure 6. Histogram of latitude.

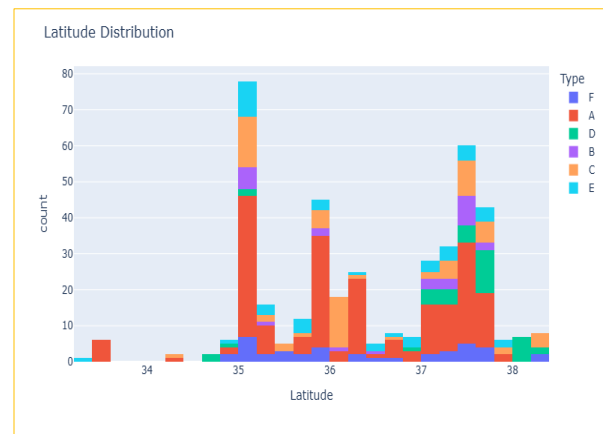


Figure 7. Segmented histogram of latitude.

According to above histograms of latitude, several observations are noteworthy. When latitude is below 34.5, most of sinkholes are stemmed from water supply pipe damages, while the proportions of this type are high along three peaks of 35, 35.9 and 37.5. Most of all, this proportion is highest along latitude of 35.9. Three modes of latitude correspond to the locations of three metropolitan areas, playing a role of one axis of clusters.

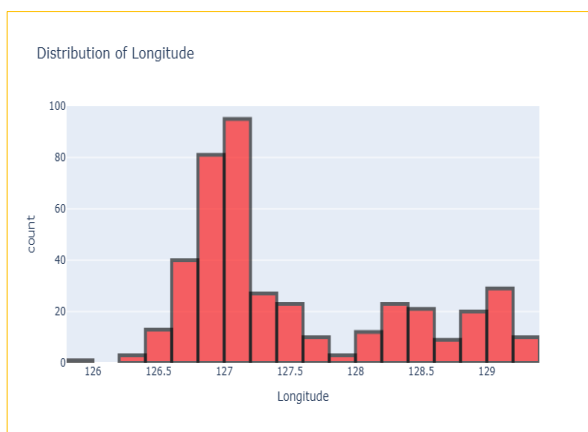


Figure 8. Histogram of longitude.

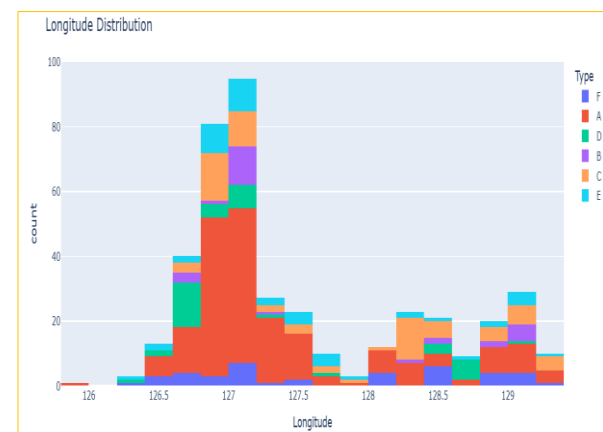


Figure 9. Segmented histogram of longitude.

Subsequent longitude distribution reveals dominance of water supply pipe damage type around latitude of 127. Like the case of latitude, two or three modes might correspond to the longitudes of two or three metropolitan areas.

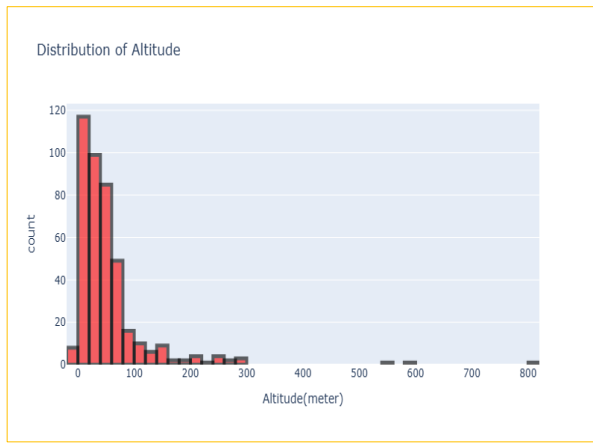


Figure 10. Histogram of altitude.

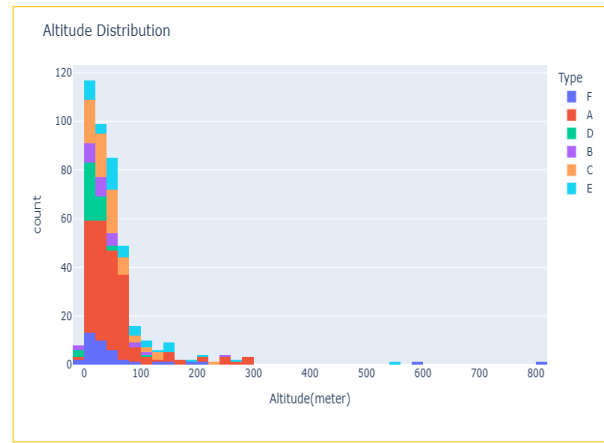


Figure 11. Segmented histogram of altitude.

Distribution of altitude is strongly skewed to the right with 3 outliers on the right side (Figure 10 and Figure 11). The mean altitude is 53 meters while 66 percent of sinkholes are recorded below 50 meters. Sinkholes caused by poor excavation during a construction procedure (Type D) do not appear above 400 meters, which seems reasonable since constructions are limited at such a high altitude.

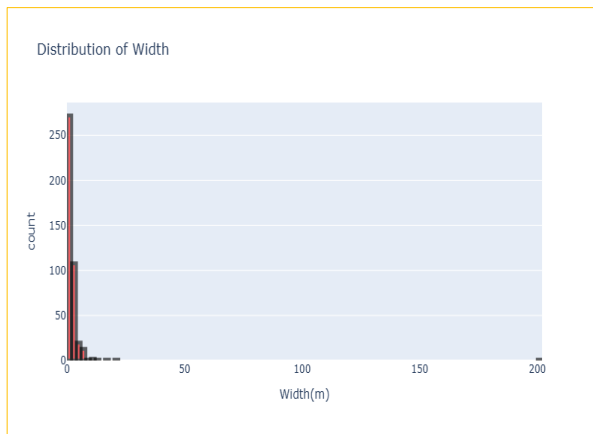


Figure 12. Histogram of width.

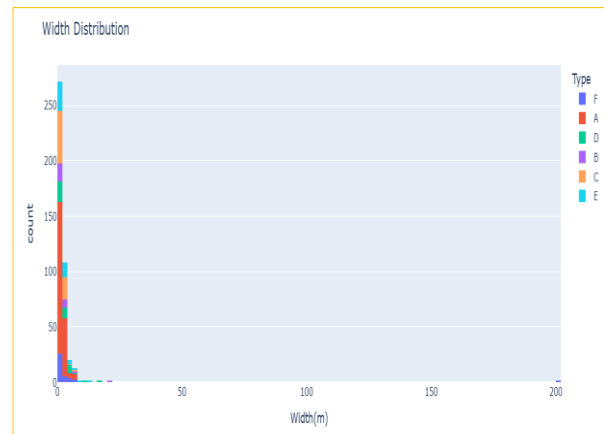


Figure 13. Segmented histogram of width.

Width measure of sinkhole disaster is a key feature in a classification issue. The mean width is 2.23 meters, while 92 percent of sinkholes are less than 3 meters wide. Remarkable fact is that all sinkholes measuring more than 10 meters in width are caused by poor excavations or underground facilities damages, implying that more serious sinkholes can appear when infrastructure construction management is inadequate rather than when supply or drain water pipe is outdated or risky.

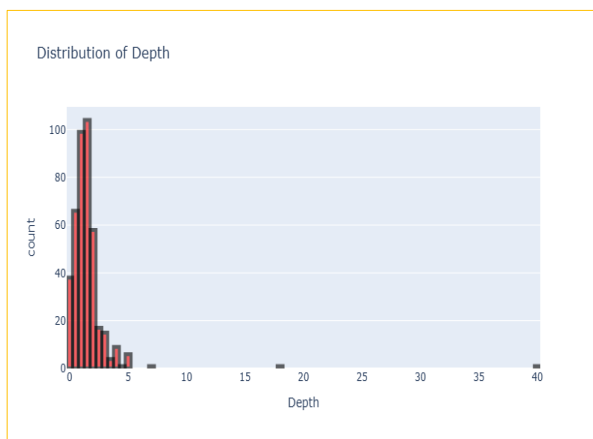


Figure 14. Histogram of depth.

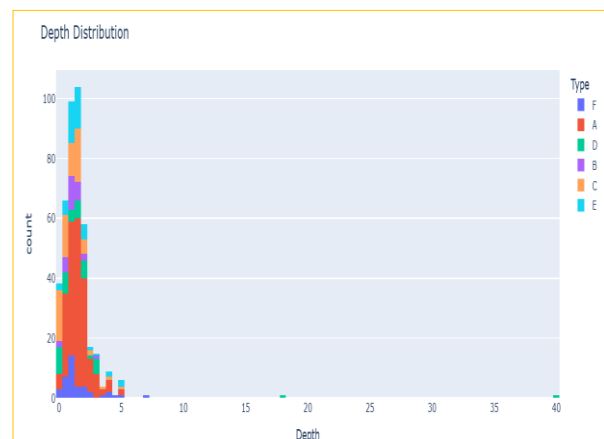


Figure 15. Segmented histogram of depth.

Distributions of depth and width are similar in terms of severe skewness and the existence of outliers. Like the variable of width, for the sinkholes deeper than 3 meters, the proportion of underground facilities

damage significantly increases, where several outliers belong to this category. Graphical displays of these variables in this section suggest that several underlying patterns can be identified and encompassed in the subsequent statistical and machine learning models. The distributions of other variables such as number of casualties, number of deaths, number of car damages, and construction permission rate are not included in this section, which will be covered later. Among these, yearly construction permission rate is not directly matching to each sinkhole record but to the cities where some of sinkhole events are simultaneously included.

First Data Inspection: Linear Regression Model

As a first step to figure out the relations of variables, multiple linear regression with a backward selection is conducted after encoding sinkhole types into integer values from 0 to 5. Among 11 explanatory variables, 6 variables are selected with a standard of p-values less than 0.1. Estimated prediction model with standard errors in parenthesis is:

$$\text{Sinkhole type number} = -41.557_{(14.193)} + 0.139_{(0.08)} * \text{Latitude} + 0.308_{(0.104)} * \text{Longitude} + 0.016_{(0.009)} * \text{Width} + 0.030_{(0.014)} * \text{Width extension} + 3.564_{(1.228)} * \text{Death} - 0.064_{(0.034)} * \text{Month}$$

However, adjusted R^2 is only 0.062, implying explanatory power is too low. Output table is given in Appendix as Table 6 and Table 7. To back up the low explanatory power, more features outside the data sets from Underground Safety Information Network are needed.

Data Extension

The variables from Underground Safety Information Network are designating the status of sinkhole events, describing some distinct features about sinkholes. Six classification standards in Korean sinkholes are not symptom based but causality based, established to make it fast and easy to restore. Therefore, in terms of the susceptibility of sinkholes, several key features such as the ages of water supply and drainpipes or proximity to large-scale construction sites are missing, most of which are hard to be obtained during the data collection stages and thus will be included in next study. Leveraging these limitations, underground geological types are added in our study, since geological characteristics are more closely related to susceptibility of sinkhole attacks in terms that even the same risk of supply or drainpipe damage or excavation status might be differently realized if the underlying geological types are different. Therefore, underlying geological types are gathered from Geo Big Data Platform (<https://data.kigam.re.kr/>), as a proxy for missing key features in one sense and as a way of increasing explanatory power in the other sense. These underlying geological features are searched by one-by-one matching of inserting each address of sinkhole attacks into the search engine of Geo Big Data Platform and then classified into 21 subgroups as described in Table 4.

Table 4. Geological classes underlying sinkhole regions.

Notation of geological class	Name
G	Gneiss
H	Deabo granite
I	Bokcheon granite
J	Yucheon Group
K	Northern Sangwon Supergroup
L	Hayang Group
M	Nangrim Group
N	Sindong Conglomerate
P	Basaltic granite
Q	Basaltic trachyandesite
R	Foliated granite
S	Buncheon granitic rocks
T	Yangdeoktong (Yangdoktong) Formation
U	Hamgyeong Formation
V	Okcheon Supergroup
W	Duman Formation
X	Granitic gneiss
Y	Namsan granitic rocks
Z	Hamdeok Group
ZZ	Daedong Supergroup
ZW	Reclaimed land

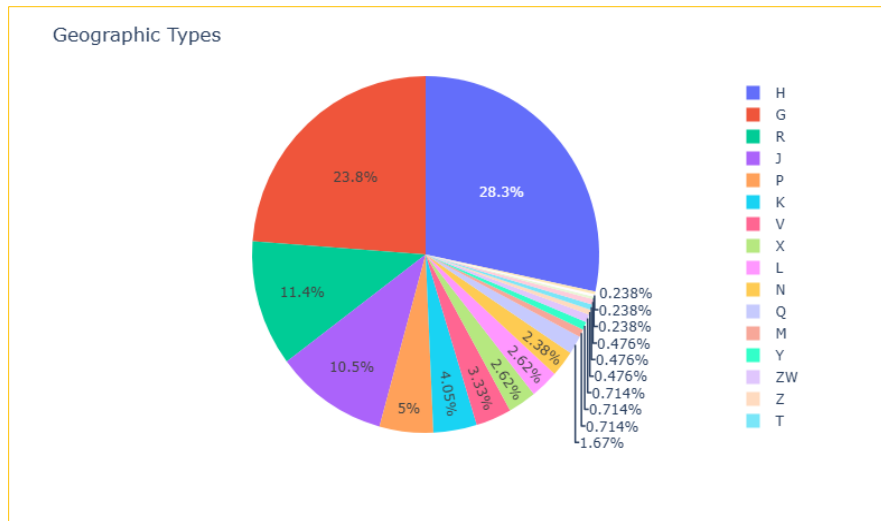


Figure 16. Pie chart of geological classes.

Below the sinkhole attack regions, the most common geological classes, such as G (Gneiss), H (Deabo granite), R (Foliated granite) and J (Yucheon Group) composed 74% of total sinkholes, differing from the proportions of all regions with or without sinkhole attacks. Therefore, certain type of dependencies can be presumed between geological types and sinkhole attacks, which is apparent in geotechnical context.

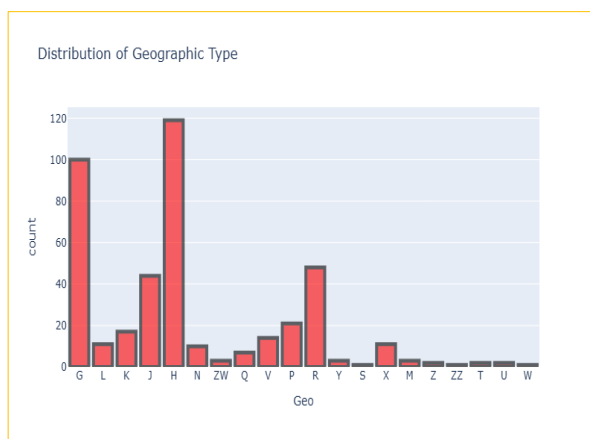


Figure 17. Bar chart of geological classes.

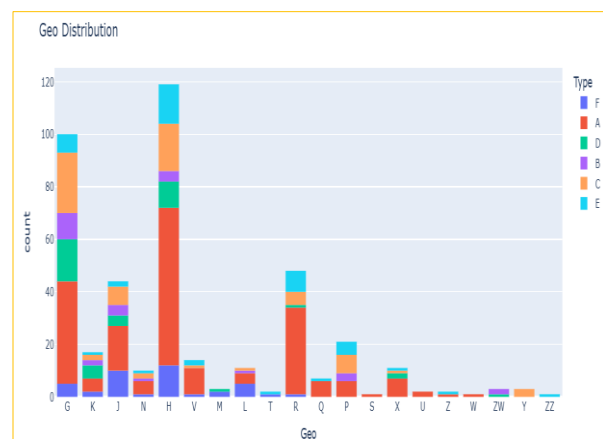


Figure 18. Segmented bar chart of geological classes.

More specifically, from the segmented histogram, sinkholes from poor excavation and drain-pipe damage are almost negligible in geographic type R (Foliated granite). Also, water supply pipe damage is not a cause in the regions of M (Nangrim Group), T (Yangdeoktong Formation), ZW (Reclaimed land), Y (Namsan granitic rocks), and ZZ (Daedong Supergroup).

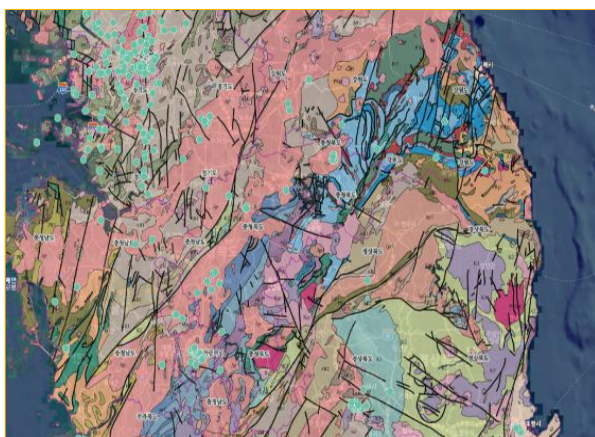


Figure 19. Map of sinkhole and geology (I).



Figure 20. Map of sinkhole and geology (II).

From above two maps with sinkhole attack regions (green circle) and geological types below these districts, several clusters, one north-western and the other south-eastern are noticeable. To catch up with more thorough relations, second data investigation is included in the next section.

Second Data Investigation: Factor Analysis

To more closely look into the relations of these 13 variables, Principal Component Analysis (PCA) is implemented after encoding geographical types into integer values from 1 to 21 and sinkhole types into the numbers from 0 to 5, minimizing the issues related with the application of PCA in case of categorical variables. Factor loadings are in Table 9 of appendix.

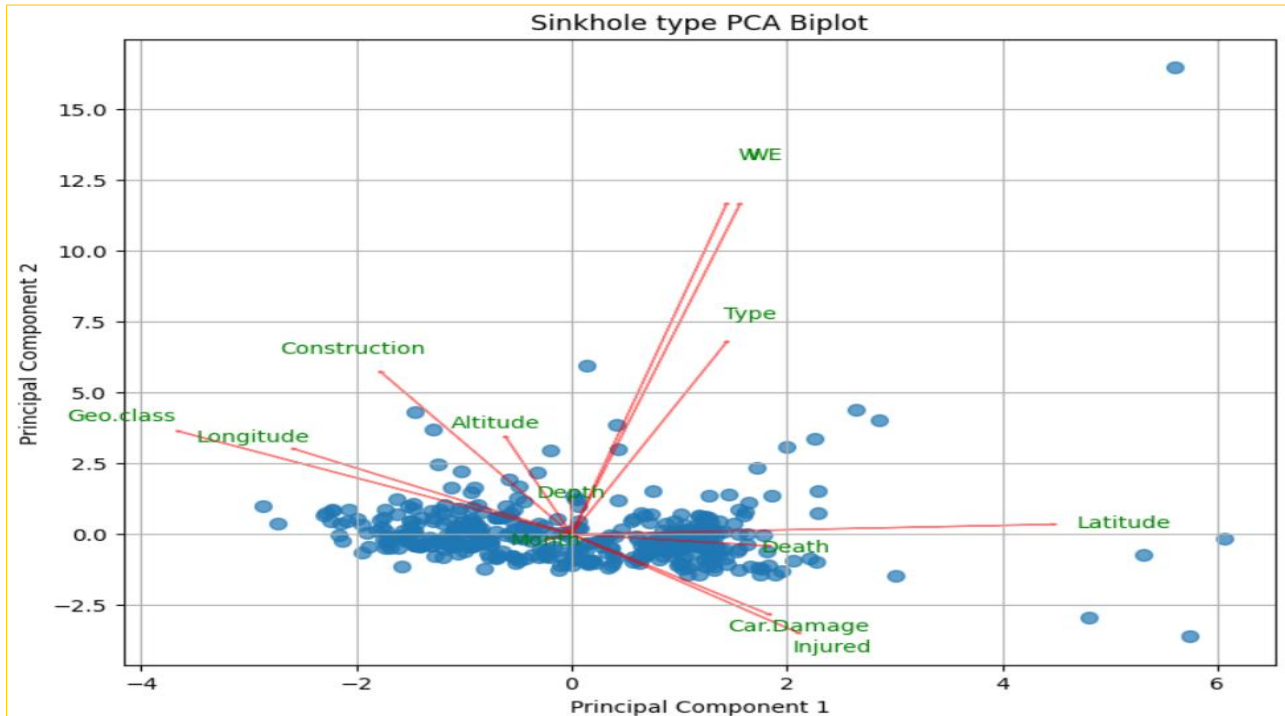


Figure 21. PCA biplot of sinkhole features.

In Figure 21, 'width' and 'width extensions' are denoted as 'W' and 'WE' for simplicity. The negative correlation between the pair of 'longitude' and 'geological classes' and that of 'latitude' and 'death' composes the most of the first principal component. The positive correlation between 'width' and 'sinkhole type' makes up the most of the second principal component. Moreover, sinkhole type is positively correlated with 'width and width extensions', 'altitude', 'construction permission rate' and 'latitude', while it is almost uncorrelated with 'number of car damages', 'number of injured', 'longitude' and 'geo class'. In terms of correlation, width measure of sinkhole attacks is a key variable determining sinkhole classification. Geological class does not directly govern the variation of sinkhole types; however, it implicitly but strongly affects all variations of all 13 features. These underlying dependencies will be utilized in the following Random Forest Classification method.

Final Data Inspection: Outcome of Random Forest Classifier

Leveraging decision trees while enhancing their accuracy, Random Forest (RF) stands out as a highly effective ML method employed for both regression and classification purposes [23].

To assess classifier performance, several measures are applied:

True Positives (TP): instances correctly labeled as positive.

True Negatives (TN): the instances correctly labeled as negative.

False Positives (FP): instances mistakenly labeled as positive.

False Negatives (FN): instances mistakenly labeled as negative.

These measures are computed using a confusion matrix, a foundation for evaluating performance. For a single class C_i , the terms TP_i , FN_i , TN_i , and FP_i , are used to evaluate class-specific metrics. With these measures, the followings are calculated to evaluate the performance:

$$\text{Sensitivity} = \frac{TP_i}{(TP_i + FN_i)}$$

$$\text{Specificity} = \frac{TN_i}{(TN_i + FP_i)}$$

$$\text{Accuracy} = \frac{(TP_i + TN_i)}{(TN_i + FP_i + TP_i + TN_i)}$$

$$\text{Precision} = \frac{TP_i}{(TP_i + FP_i)}$$

$$\text{F1-score} = \frac{2 * \text{Precision} * \text{Sensitivity}}{(\text{Precision} + \text{Sensitivity})}$$

In our study, classification report exhibits accuracy rate of 94.7% as in the Table 5.

Table 5. Classification report of random forest classifier.

Classification Report:				
	precision	recall	f1-score	support
class 0	1.00	1.00	1.00	4
class 1	1.00	0.75	0.86	4
class 2	0.89	1.00	0.94	8
class 3	1.00	1.00	1.00	1
class 4	1.00	1.00	1.00	1
class 5	1.00	1.00	1.00	1
accuracy			0.95	19
macro avg	0.98	0.96	0.97	19
weighted avg	0.95	0.95	0.95	19

Associated confusion matrix is given in appendix as Table 8 where diagonal elements are 4, 3, 8, 1, 1, and 1, allowing high accuracy, recall and F1-scores.

Conclusion

To identify and classify sinkholes, data sets such as LiDAR or images from satellites or drones are now used more and more widely. In our model, only open access field data like address, width, depth, month, underlying geological classes of sinkholes and so on are used to classify sinkhole types, obtaining 94.7% accuracy rate, along with 98% precision and 96% recall rate. If the data sets are extended to include thermal data and the depth of underground water below the sinkhole attacks as are attached in appendix, which will be included in our next study, more accurate classification outcome is expected. Moreover, if data sets from various sensors are merged into this background field data, earlier warning and identification of sinkhole attacks seem to become more plausible, which is a policy proposal we want to address.

Appendix

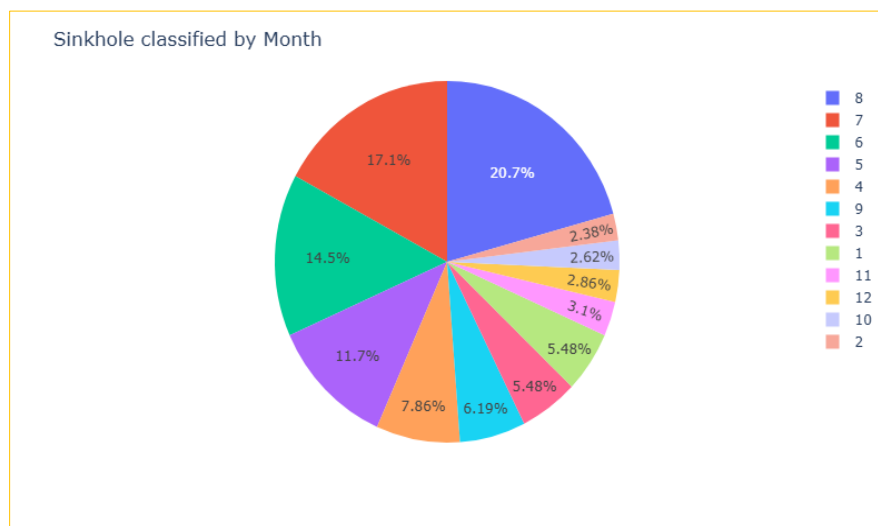


Figure 22. Pie chart of month.

Table 6. Regression outcome (I).

Linear Regression ▼

Model Summary - Type Number

Model	R	R ²	Adjusted R ²	RMSE
M ₀	0.285	0.081	0.059	1.729
M ₁	0.285	0.081	0.061	1.727
M ₂	0.283	0.080	0.062	1.726
M ₃	0.279	0.078	0.062	1.726
M ₄	0.275	0.076	0.062	1.726

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
M ₀	Regression	108.268	10	10.827	3.621	< .001
	Residual	1222.922	409	2.990		
	Total	1331.190	419			
M ₁	Regression	108.048	9	12.005	4.024	< .001
	Residual	1223.142	410	2.983		
	Total	1331.190	419			
M ₂	Regression	106.350	8	13.294	4.461	< .001
	Residual	1224.841	411	2.980		
	Total	1331.190	419			
M ₃	Regression	103.857	7	14.837	4.981	< .001
	Residual	1227.333	412	2.979		
	Total	1331.190	419			
M ₄	Regression	101.024	6	16.837	5.653	< .001
	Residual	1230.167	413	2.979		
	Total	1331.190	419			

Table 7. Regression outcome (II).

M ₃	(Intercept)	-40.912	14.210		-2.879	0.004
	Latitude	0.140	0.080	0.086	1.756	0.080
	Longitude	0.303	0.105	0.141	2.897	0.004
	Width(m)	0.016	0.009	0.088	1.719	0.086
	Width.extention	0.030	0.014	0.107	2.096	0.037
	Depth	0.036	0.037	0.046	0.975	0.330
	Death	3.564	1.228	0.138	2.902	0.004
	Month	-0.066	0.034	-0.094	-1.961	0.051
M ₄	(Intercept)	-41.557	14.193		-2.928	0.004
	Latitude	0.139	0.080	0.086	1.752	0.080
	Longitude	0.308	0.104	0.143	2.953	0.003
	Width(m)	0.016	0.009	0.086	1.683	0.093
	Width.extention	0.030	0.014	0.109	2.137	0.033
	Death	3.564	1.228	0.138	2.902	0.004
	Month	-0.064	0.034	-0.090	-1.887	0.060

Table 8. Confusion matrix.

```
[ [4 0 0 0 0 0]
  [0 3 1 0 0 0]
  [0 0 8 0 0 0]
  [0 0 0 1 0 0]
  [0 0 0 0 1 0]
  [0 0 0 0 0 1] ]
```

Table 9. Factor loadings of PCA.

array([[	0.73590015,	0.02102047],
[	-0.42408168,	0.18205483],
[	0.23756615,	0.70677843],
[	0.2579248,	0.7062594],
[	-0.00090089,	0.07679025],
[	0.29921979,	-0.02470978],
[	0.34721914,	-0.2108418],
[	0.30190563,	-0.17109033],
[	-0.10062144,	0.20745991],
[	-0.03514182,	-0.01092037],
[	-0.29172323,	0.34642501],
[	-0.60093702,	0.21977311],
[	0.23747413,	0.41105175]])

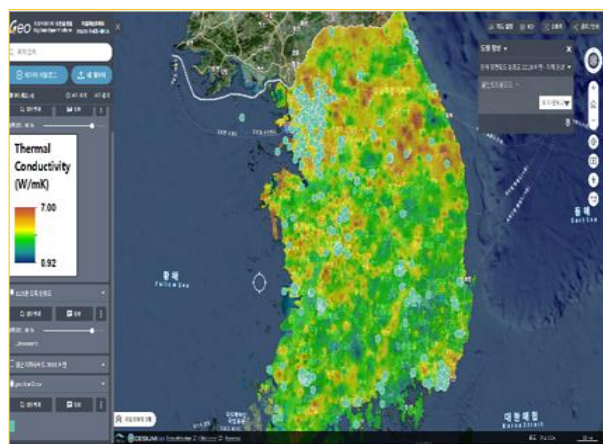


Figure 23. Map of thermal data.

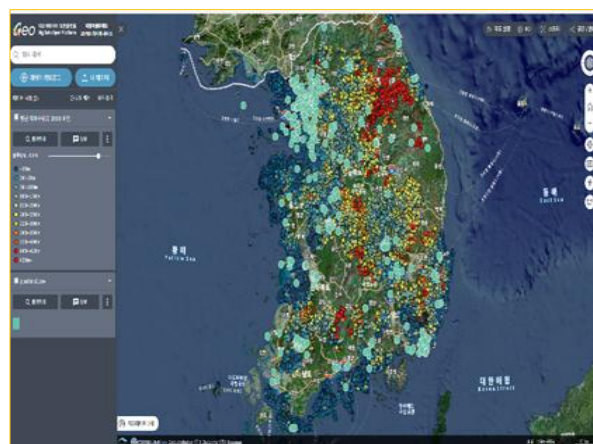


Figure 24. Map of underground water depth.

Declarations

Acknowledgments: Authors are thankful to the entire staff of Seoul Innovations Research Institute for insightful ideas and invaluable proofreading.

Author Contributions: WJC: Definition of intellectual content, data collection, prepared first draft of manuscript, review manuscript, data analysis; CJC: Design of study, statistical analysis and interpretation, data collection, manuscript preparation, editing, prepared first draft of manuscript; HC: Literature survey, prepared first draft of manuscript, data collection, data analysis, python coding; SC: Literature survey, data collection, data analysis, manuscript revision, python coding; DYH: Manuscript preparation, data collection, manuscript review, statistical analysis; MJ: Concept, design of study, manuscript preparation, data collection, data analysis, prepared first draft of manuscript, manuscript revision; BN: Manuscript review, statistical interpretation.

Conflict of Interest: The authors declare no conflict of interest.

Consent to Publish: The authors agree to publish the paper in International Journal of Recent Innovations in Academic Research.

Data Availability Statement: The datasets generated and/or analyzed during this study are not publicly available but are available from the corresponding author upon reasonable request.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Research Content: The research content of manuscript is original and has not been published elsewhere.

References

- Hasan, M.Z., Hanapi, Z.M., Zukarnain, Z.A., Huyop, F.H. and Abdullah, M.D.H. 2025. An efficient detection of sinkhole attacks using machine learning: Impact on energy and security. *PloS One*, 20(3): e0309532.
- Yavariabdi, A., Kusetogullari, H., Orhan, O., Uray, E., Demir, V., Celik, T. and Mendi, E. 2023. SinkholeNet: A novel RGB-slope sinkhole dataset and deep weakly-supervised learning framework for sinkhole classification and localization. *The Egyptian Journal of Remote Sensing and Space Sciences*, 26(4): 966-973.
- Kariminejad, N., Shahabi, H., Ghorbanzadeh, O., Shafaie, V., Hosseinalizadeh, M., Homayouni, S. and Pourghasemi, H.R. 2024. Evaluation of various deep learning algorithms for landslide and sinkhole detection from UAV imagery in a semi-arid environment. *Earth Systems and Environment*, 8(4): 1387-1398.
- Kulshrestha, A. 2023. InSAR time series analysis for sinkhole detection using deep learning. Ph.D. Thesis, University of Twente, Faculty of Geo-Information Science and Earth Observation (ITC). <http://doi.org/10.3990/1.9789036557283>
- Bouali, M., Ababsa, F., Sammuneh, M.A., El Meouche, R., Salavati, B. and Viguier, F. 2025. Enhanced railway sinkhole detection and characterization using lidar-based 3d modeling and geometric analysis. *Pattern Recognition Letters*, 197: 378-384.
- Jiang, Z., Hu, S., Deng, H., Wang, N., Zhang, F., et al. 2024. Detection and automatic identification of loess sinkholes from the perspective of LiDAR point clouds and deep learning algorithm. *Geomorphology*, 465: 109404.

7. Muili, O. and Babaie, H.A. 2025. Sinkhole susceptibility analysis using machine learning for west central Florida. *Applied Computing and Geosciences*, 27: 100262.
8. Lee, E.J., Shin, S.Y., Ko, B.C. and Chang, C. 2016. Early sinkhole detection using a drone-based thermal camera and image processing. *Infrared Physics and Technology*, 78: 223-232.
9. Zhu, J. and Pierskalla Jr, W.P. 2016. Applying a weighted random forests method to extract karst sinkholes from LiDAR data. *Journal of Hydrology*, 533: 343-352.
10. Kang, M.S., Kim, N., Im, S.B., Lee, J.J. and An, Y.K. 2019. 3D GPR image-based U-Net for enhancing underground cavity detectability. *Remote Sensing*, 11(21): 2545.
11. Mihevc, A. and Mihevc, R. 2021. Morphological characteristics and distribution of dolines in Slovenia, a study of a lidar-based doline map of Slovenia. *Acta Carsologica*, 50(1): 11-36.
12. Nefeslioglu, H.A., Tavus, B., Er, M., Ertugrul, G., Ozdemir, A., Kaya, A. and Kocaman, S. 2021. Integration of an InSAR and ANN for sinkhole susceptibility mapping: A case study from Kirikkale-Delice (Turkey). *ISPRS International Journal of Geo-Information*, 10(3): 119.
13. Rafique, M.U., Zhu, J. and Jacobs, N. 2022. Automatic segmentation of sinkholes using a convolutional neural network. *Earth and Space Science*, 9(2): e2021EA002195.
14. Alrabayah, O., Caus, D., Watson, R.A., Schulten, H.Z., Weigel, T., Rüpke, L. and Al-Halbouni, D. 2024. Deep-learning-based automatic sinkhole recognition: Application to the eastern dead sea. *Remote Sensing*, 16(13): 2264.
15. DeWaele, J. and Gutiérrez, F. 2022. Karst hydrogeology, geomorphology and caves. John Wiley and Sons: Hoboken, NJ, USA, 1-912pp.
16. Gutiérrez, F., Parise, M., De Waele, J. and Jourde, H. 2014. A review on natural and human-induced geohazards and impacts in karst. *Earth-Science Reviews*, 138: 61-88.
17. Perrin, J., Cartannaz, C., Noury, G. and Vanoudheusden, E. 2015. A multicriteria approach to karst subsidence hazard mapping supported by weights-of-evidence analysis. *Engineering Geology*, 197: 296-305.
18. Wang, G., Hao, J., Wen, H. and Cao, C. 2022. A random forest model of karst ground collapse susceptibility based on factor and parameter coupling optimization. *Geocarto International*, 37(27): 15548-15567.
19. Xie, Y.H., Zhang, B.H., Liu, Y.X., Liu, B.C., Zhang, C.F. and Lin, Y.S. 2022. Evaluation of the karst collapse susceptibility of subgrade based on the AHP method of ArcGIS and prevention measures: A case study of the Quannan expressway, section K1379+ 300-K1471+ 920. *Water*, 14(9): 1432.
20. Galve, J.P., Gutiérrez, F., Remondo, J., Bonachea, J., Lucha, P. and Cendrero, A. 2009. Evaluating and comparing methods of sinkhole susceptibility mapping in the Ebro Valley evaporite karst (NE Spain). *Geomorphology*, 111(3-4): 160-172.
21. Taheri, K., Shahabi, H., Chapi, K., Shirzadi, A., Gutiérrez, F. and Khosravi, K. 2019. Sinkhole susceptibility mapping: A comparison between bayes-based machine learning algorithms. *Land Degradation and Development*, 30(7): 730-745.
22. Bilgilioglu, S.S., Gezin, C., Iban, M.C., Bilgilioglu, H., Gündüz, H.I. and Arslan, Ş. 2025. Explainable sinkhole susceptibility mapping using machine-learning-based SHAP: Quantifying and comparing the effects of contributing factors in Konya, Türkiye. *Applied Sciences*, 15(6): 3139.
23. Breiman, L. 2001. Random forests. *Machine Learning*, 45: 5-32.
24. Elmahdy, S.I., Mohamed, M.M., Ali, T.A., Abdalla, J.E.D. and Abouleish, M. 2022. Land subsidence and sinkholes susceptibility mapping and analysis using random forest and frequency ratio models in Al Ain, UAE. *Geocarto International*, 37(1): 315-331.
25. Yilmaz, I. 2007. GIS based susceptibility mapping of karst depression in gypsum: A case study from Sivas basin (Turkey). *Engineering Geology*, 90(1-2): 89-103.
26. Kim, Y.J., Nam, B.H., Jung, Y.H., Liu, X., Choi, S., Kim, D. and Kim, S. 2022. Probabilistic spatial susceptibility modeling of carbonate karst sinkhole. *Engineering Geology*, 306: 106728.
27. Kim, K., Kim, J., Kwak, T.Y. and Chung, C.K. 2018. Logistic regression model for sinkhole susceptibility due

to damaged sewer pipes. *Natural Hazards*, 93(2): 765-785.

28. Zhou, G., Yan, H., Chen, K. and Zhang, R. 2016. Spatial analysis for susceptibility of second-time karst sinkholes: A case study of Jili Village in Guangxi, China. *Computers and Geosciences*, 89: 144-160.
29. Mohammady, M., Pourghasemi, H.R., Amiri, M. and Tiefenbacher, J.P. 2021. Spatial modeling of susceptibility to subsidence using machine learning techniques. *Stochastic Environmental Research and Risk Assessment*, 35(8): 1689-1700.
30. Ozdemir, A. 2015. Investigation of sinkholes spatial distribution using the weights of evidence method and GIS in the vicinity of Karapinar (Konya, Turkey). *Geomorphology*, 245: 40-50.
31. Krizhevsky, A., Sutskever, I. and Hinton, G.E. 2012. Imagenet classification with deep convolutional neural networks. In: Pereira, F., Burges, C.J.C., Bottou, L. and Weinberger, K.Q., (Eds.), *Advances in neural information processing systems 25 (NIPS 2012)*. Curran Associates, Inc. 1097–1105p.

Citation: Woohyun James Chun, Chang Joon Chung, Hyungjun Chin, Seongbin Cho, Dong Yeon Hu, Madeline Jung and Bob Nam. 2025. Automatic Sinkhole Classification by Employing Field Data: Application to Sinkholes in South Korea. *International Journal of Recent Innovations in Academic Research*, 9(4): 24-37.

Copyright: ©2025 Woohyun James Chun, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.