Geomechanics and Engineering, *Vol. 41, No. 1 (2025) 11-19*
https://doi.org/10.12989/gae.2025.41.1.011

Machine learning-based prediction of underground utility counts using electrical resistance numerical data

Hee-Hwan Ryu[1a], Suyoung Choi[2b], Song-Hun Chong[3c], Tae-Young Kim[3d], Jiyun Lee[1e] and Meiyan Kang*[2]

[1]*Transmission & Substation Laboratory, Korea Electric Power Corporation (KEPCO) Research Institute, Daejeon 34056, Republic of Korea*
[2]*Department of Mathematics, Ajou University, 206 World cup-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 16499, Republic of Korea*
[3]*Department of Civil Engineering, Sunchon National University, 255 Jungang-ro, Sunchon, Jeollanam-do 57922, Republic of Korea*

(*Received December 7, 2024, Revised February 27, 2025, Accepted March 5, 2025*)

Abstract. Understanding subsurface conditions is essential for mitigating unexpected hazards during excavation, particularly in areas with underground utilities. Electrical resistance values play a crucial role for predicting these conditions. This study develops a non-destructive, cost-effective framework to predict the number of buried pipelines using machine learning applied to numerical electrical resistance data. The resistance data was generated by applying a numerical electrical resistance model, developed using generalized mesh techniques, based on electrode and structural geometric parameters. A total of 87,572 data samples, comprising 56 electrodes across 667 cases and 90 electrodes across 558 cases, were used. Various machine learning techniques, including Support Vector Machine, Random Forest, and Extreme Gradient Boosting, were employed to classify underground utility counts. Additionally, a deep learning method, specifically Convolutional Neural Network, transformed the resistivity data into a 2D matrix format for analysis. The results indicate that the data provides sufficient information to accurately determine the number of buried pipelines, demonstrating the potential of these models for underground utility prediction. This work integrates numerical simulations with machine learning to develop a model capable of underground utility prediction. Given the significant challenges associated with collecting and processing real-world data for such applications, utilizing simulation data is essential to demonstrate the feasibility of these models.

Keywords: convolutional neural networks; electricity resistance survey; machine learning; underground utility detection

1. Introduction

Underground utility detection is essential in modern construction and maintenance, as it ensures the safety and integrity of infrastructure projects (Uslu *et al*. 2014, Lew 1996). Accurately identifying subsurface conditions help prevent hazardous incidents, such as damage to underground utilities during construction activities (Jeong, Abraham and Lew 2004). Such damage can result in service disruptions, costly repairs, and potential legal liabilities. For instance, hitting a gas line can cause explosions, while damaging water pipelines can lead to flooding and property damage. Therefore, effective subsurface utility detection is crucial for mitigating these risks and ensuring the successful completion of construction projects (Lew 2020). The ability to detect underground utilities minimizes project delays, reduces unexpected costs, and ensures the safety of construction workers and the public.

*Corresponding author, Ph.D. Candidate
E-mail: miyeon@ajou.ac.kr
[a]Ph.D.
[b]Ph.D.
[c]Ph.D.
[d]Graduate Student
[e]Graduate Student

In recent years, non-destructive methods for detecting underground utilities have become widely adopted due to their efficiency and minimal damage to buried objects (Reiter *et al.* 2023). Commonly employed techniques include electrical resistivity surveys, Ground Penetrating Radar (GPR), and impact-echo. GPR uses electromagnetic waves to create subsurface images, making it effective for detecting a wide range of materials. However, while GPR is highly effective in dry, sandy soils, its effectiveness diminishes in clay-rich soils due to signal attenuation (Tosti *et al.* 2016). In contrast, the impact-echo method utilizes stress waves generated by mechanical impacts to detect flaws and features within concrete structures (Zembillas 2003). This method excels at detecting voids and cracks in concrete but may not provide detailed images of deeper utilities (Yehia *et al.* 2007, Lee *et al.* 2024).

Electrical resistivity surveys stand out for their ability to assess subsurface conditions accurately across diverse soil types (Samouëlian *et al*. 2005, Olabode and San 2023). Unlike GPR and impact-echo, which are influenced by factors such as material composition and environmental conditions, the electrical resistivity survey is less affected by these variables, making it particularly effective for mapping underground features (Loke *et al.* 2021). This method measures soil resistance to electrical current by injecting current through electrodes and recording the resulting voltage differences, which are used to infer subsurface resistivity distributions. Distinct materials,

ISSN: 2005-307X (Print), 2092-6219 (Online)

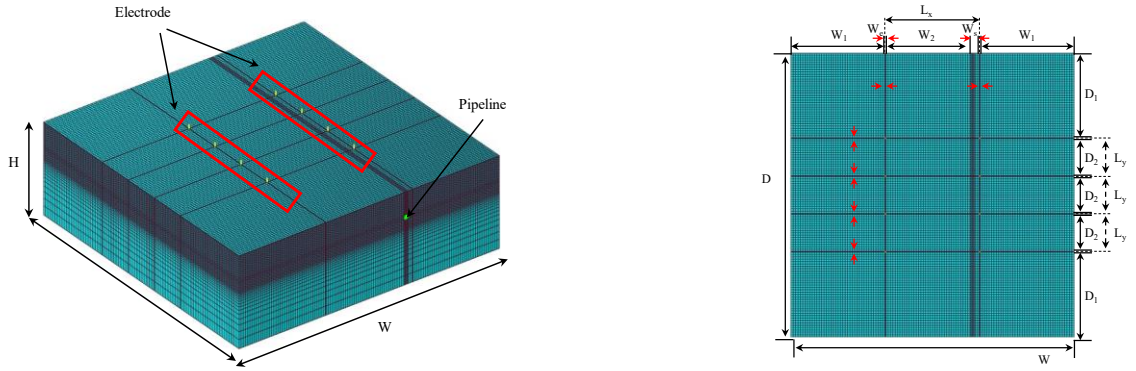
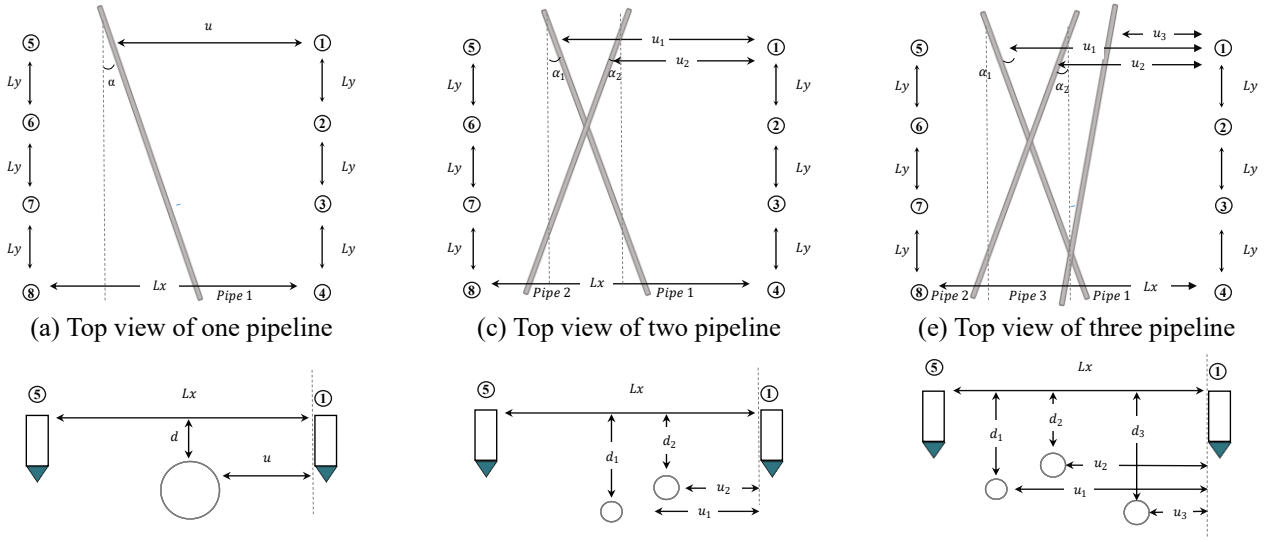


Fig. 1 Electrical resistance numerical module



(a) Top view of one pipeline (c) Top view of two pipeline (e) Top view of three pipeline

(b) Side view of one pipeline (d) Side view of two pipelines (f) Top view of three pipelines

Fig. 2 Top and side views of the simulation under three conditions: one pipeline buried, two pipelines buried, and three pipelines buried.

including soil, rock, and buried utilities, exhibit unique resistivity values, enabling accurate identification of underground features (Herman 2001, Chhun and Yune 2023). Our analysis utilized numerically analyzed resistance data generated with COMSOL® (2023), a reliable and consistent dataset validated by Kim *et al*. (2024). Results derived from this simulated dataset indicate that electrical resistivity data contains valuable information regarding underground utilities, highlighting their potential for further in-depth research.

The rapid advancement of machine learning has led to increasing integration into geotechnical engineering (Tariq *et al.* 2024, Samadi *et al.* 2024), particularly for underground utility detection, by combining traditional methods such as GPR, impact-echo, and electrical resistivity surveys. These techniques enhance accuracy and reliability by identifying patterns and correlations in data that may not be apparent through conventional analysis methods, leading to more accurate predictions. Machine learning algorithms, such as Support Vector Machines (SVM), Random Forest, and Extreme Gradient Boosting (XG-Boost), are well-suited for classification tasks and have been successfully applied to predict the presence and characteristics of underground utilities (Todkar *et al.* 2019, Jaufer *et al.* 2022). Furthermore, Convolutional Neural Networks (CNNs), combined with the impact-echo method and time-frequency images from wavelet transforms, have been effectively used to detect voids behind tunnel linings (Lee *et al*. 2024). GPR imaging integrated with the YOLO deep learning model has also been employed to detect subsurface cracks (Abdelmawla *et al*. 2023).

Despite the growing body of research on underground utility detection, no studies have explicitly focused on predicting the count of underground utilities which is a critical parameter for risk-informed excavation planning and resource allocation. This study addresses this gap by utilizing electrical resistivity surveys with numerical simulation to develop machine leaning models capable of underground utility prediction.

Given the significant challenges associated with collecting and processing real-world data, which is affected by various interference factors such as soil composition, rock formations, and environmental noise, simulation-based research is essential for validating research feasibility. To explore this potential, this study employs three machine learning techniques—SVM, Random Forest, and XG-

Table 1 Notation

Notation	Description	Count	Range	Unit
$R_{i,j}$	The resistance value between ground electrode ⓘ and electrode ⓙ.	56	0–11,390.0	Ω
L_x	The width of the road.	1	3.5–20	M
L_y	The distance between electrode ⓘ and electrode ⓙ.	6	8	M
α	The angle formed between the direction of a single pipe and the	1–3	0–20	o
u	The horizontal distance on the x-axis between the pipe and the first	1–3	0.3–15.0	M
d	The depth at which the pipe is buried.	1–3	0.3–10.0	m

Fig. 3 Data split for the training and test sets

Boost—along with one deep learning method, CNN, to analyze numerical simulation data and assess the applicability of electrical resistivity surveys for predicting the number of buried utilities. All the models give similar results; however, CNN shows slightly higher accuracy rather than traditional machine learning techniques. This is attributed not only to the advanced capabilities of deep learning models but also to the sophisticated feature extraction processes involved in data preprocessing.

The rest of this paper is structured as follows: Chapter 2 introduces the dataset we used; Chapter 3 presents the machine learning processes and results; and Chapter 4 discusses the additional feature extraction methods and deep learning techniques applied to enhance predictive accuracy.

2. Simulation data from numerical analysis

This study utilized a resistance dataset generated through modeling and numerical analysis using COMSOL® (2023), a finite element software program.

A numerical analysis model was developed by employing a generalized 3D mesh technique for pipelines, electrodes, and their surroundings, created using mapped and swept functions (Kim *et al.* 2024, Hong *et al.* 2022). This mesh technique produced relatively accurate electrical resistance values compared to the analytical solution for a cylindrical electrode with a conical tip. The error between the electrical resistance values obtained from the theoretical equation and numerical simulation is calculated as follows

$$ER = \frac{|R^{Theory} - R^{Num}|}{R^{Theory}} \quad (1)$$

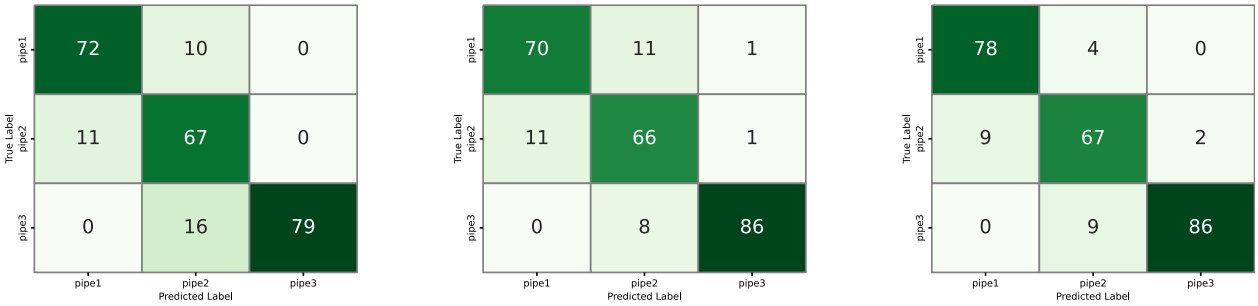
Where R^{Theory} represents the theoretical values and R^{Num} denotes the numerical analyzed values.

Fig. 1 shows the 3D numerical simulation module for generalizing mesh, which includes electrodes and buried pipe. The simulation involved placing four electrodes symmetrically on each side of the road along the pipe. Three scenarios were considered, corresponding to the presence of one, two, and three pipelines.

The top and side views of each case from the modules in COMSOL are illustrated in Fig. 2. The electrodes are labeled from ① to ⑧, with ⓘ and ⓙ representing the electrode labels and Ly denoting the distance between two consecutive electrodes. Additionally, u, d, and α represent the horizontal distance between the pipe and electrode ①, the burial depth, and the angle between the pipe and the road in the top view, respectively. Various burial conditions were simulated by altering Ly, u, d, and α to collect data. These notations and their ranges are summarized in Table 1.

Once the parameters were set, resistance values between arbitrary pairs of electrodes ⓘ and ⓙ were recorded. These values are denoted as $R_{i,j}$, where electrode ⓘ is grounded. This resulted in $2 \times \binom{8}{2} = 56$ resistance values for each simulation case.

A total of 87,572 data points were derived from 56 electrodes across 667 cases and 90 electrodes across 558 cases. However, only data from cases involving 56 electrodes were used for analysis. Specifically, 42,952 data points were analyzed, comprising 56 (7×8) electrodes across 252 samples for the “one pipe buried” condition, 242 samples for the “two pipelines buried” condition, and 273 samples for the “three pipelines buried” condition. Table 2 presents the statistics for resistance values.



(a) SVM (b) XG-boost (c) Radom Forest

Fig. 4 Confusion matrix for SVM, XG-boost, and Random Forest

Table 2 Statistic

-	$R_{1,2}$	$R_{1,3}$	...	$R_{8,6}$	$R_{8,7}$
Count	767	767	…	767	767
Mean	10,548	10,606	…	10,598	10,542
Std	217	198	…	193	213
Min	9,915	10,022	…	10,025	9,905
25%	10,464	10,494	…	10,485	10,462
50%	10,483	10,515	…	10,511	10,482
75%	10,615	10,683	…	10,681	10,613
Max	11,141	11,164	…	11,117	11,094

For machine learning and deep learning analyses, the data were divided into training and test datasets: the training dataset consisted of two-thirds of the total data 512 samples, while the test dataset comprised one-third of the data 255 samples, as depicted in Fig. 3. This study focuses on analyzing the electrical resistivity data to classify the number of buried pipelines present.

3. Analysis using machine learning models

In this study, we employed three commonly used machine learning algorithms—SVM, Random Forest, and XG-Boost—to perform the classification tasks.

3.1 SVM

SVM is a supervised learning algorithm widely used for classification tasks (Pisner and Schnyer 2020). It operates by identifying the optimal hyperplane that best separates data into different classes, thereby maximizing the margin between the hyperplane and the nearest data points. In this study, each data point has 56 resistance values, meaning each data point is represented in 56 dimensions. An SVM can separate such data by computing a 55-dimensional hyperplane.

To handle non-linearly separable data, we applied the SVM kernel trick (Kaestner 2013), using the Radial Basis Function as the kernel, as defined in Eq. (2).

Table 3 Classification report of the SVM model

Pipe Counts	Precision	Recall	F1-score	Support
1	0.87	0.88	0.87	82
2	0.72	0.86	0.78	78
3	1.00	0.83	0.91	95
Accuracy		0.85		255
Macro average	0.86	0.86	0.85	255
Weighted average	0.87	0.85	0.86	255

Where γ denotes the influence of a single training example, which we set to 0.01 to balance computational efficiency and model performance. Parameter C represents the regularization factor, which controls the trade-off between minimizing the training data error and maximizing the hyperplane margin. In our model, we set C to 10.

Table 3 presents the SVM model's classification performance across various pipeline conditions using key metrics: Precision, Recall, F1-score, and Support. Precision indicates the accuracy of positive predictions, while Recall measures the model's effectiveness in identifying actual positives. The F1-score provides a balanced evaluation by combining Precision and Recall, and Support shows the actual occurrences of each class, which helps assess potential class imbalances.

The SVM model achieved an overall accuracy of 85%. For the one-pipe condition, Precision and Recall values were 0.87 and 0.88, respectively, indicating strong prediction accuracy. The two-pipe condition showed lower Precision at 0.72 but a slight lower Recall of 0.86, suggesting some class overlap. The three-pipe condition yielded a perfect Precision of 1.00 but a Recall of 0.83, indicating high prediction accuracy with minor missed instances. The macro and weighted average F1-scores were 0.85 and 0.86, respectively, suggesting consistent model performance across all classes. A detailed confusion matrix is shown in Fig. 4(a).

3.2 XG-Boost

XG-Boost is an advanced ensemble learning technique

Table 4 Classification report of the XG-boost model

Pipe Counts	Precision	Recall	F1-score	Support
1	0.86	0.85	0.86	82
2	0.78	0.85	0.81	78
3	0.98	0.92	0.95	95
Accuracy		0.87		255
Macro average	0.87	0.87	0.87	255
Weighted average	0.88	0.87	0.88	255

Table 5 Classification report of the Random Forest model

Pipe Counts	Precision	Recall	F1-score	Support
1	0.90	0.95	0.92	82
2	0.84	0.86	0.85	78
3	0.98	0.91	0.94	95
Accuracy		0.91		255
Macro average	0.90	0.91	0.90	255
Weighted average	0.91	0.91	0.91	255

that sequentially builds decision trees, with each new tree addressing the errors of the previous ones. This iterative process optimizes model accuracy by minimizing the loss function through gradient descent (Ali *et al.* 2023).

The optimized XG-Boost model achieved an accuracy of 87%, with the optimal parameters identified through grid

$$\mathrm{K(a,b)} = \exp\left(-\gamma\left||a-b|\right|^2\right) \quad (2)$$

search and cross-validation (Budiman 2019). Key parameters included a learning rate of 0.1, which controlled the step size in gradient descent to prevent overfitting by slowing down the learning process. A maximum depth of 10 enabled the model to capture complex patterns, although it increased the risk of overfitting. Additionally, using 50 estimators enhanced performance by adding more trees to the ensemble, albeit with increased computational demands. These parameters were fine-tuned to balance complexity and performance, aiming for robust generalization on unseen data.

Table 4 presents the classification performance of the XG-Boost model based on Precision, Recall, F1-score, and Support. For the one-pipe condition, both Precision and Recall were 0.86, reflecting balanced accuracy in identifying this class. The two-pipe condition showed slightly lower Precision at 0.78 and Recall at 0.85, suggesting some class overlap. The three-pipe condition exhibited the best performance, with a precision of 0.98 and recall of 0.92, demonstrating high accuracy in distinguishing this class. The macro and weighted average F1-scores were 0.87 and 0.88 respectively, confirm consistent performance across all classes. A detailed confusion matrix is shown in Fig. 4(b).

3.3 Random forest

Random Forest is an ensemble learning technique that constructs multiple decision trees during training, with the final prediction determined by majority vote (classification) or average (regression) from individual trees (Parmar *et al.* 2019). This approach enhances model accuracy and robustness by reducing overfitting and improving generalization. By randomly selecting subsets of training data and features for each tree, Random Forest ensures diverse tree construction, resulting in improved performance over a single decision tree.

This model also used grid search and cross-validation to fine-tune the hyperparameters. The optimal parameters included a maximum depth of 10, which enabled the model to capture complex relationships while mitigating overfitting, and a minimum sample leaf of 1, ensuring leaf nodes contained at least one sample to capture fine-grained features. Additionally, the minimum sample split was set to 2, ensuring that nodes must have at least two samples to split, which prevents overfitting on noise. Setting n_estimators to 200 provided a balance between accuracy and computational efficiency, as additional trees can improve performance but increase computational costs.

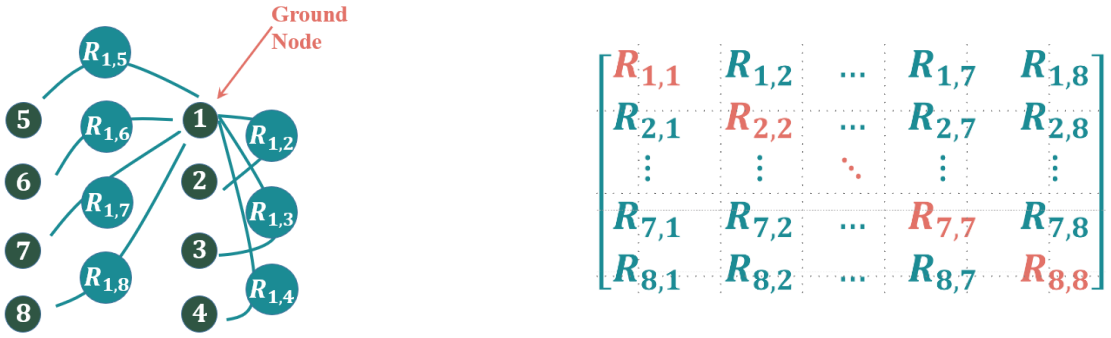
Table 5 presents the classification performance of the Random Forest model in terms of Precision, Recall, F1-score, and Support for each class. The one-pipe condition achieved high performance with a Precision of 0.90 and Recall of 0.95, indicating reliable identification accuracy. The two-pipe condition had slightly lower metrics, with a Precision of 0.84 and Recall of 0.86, showing balanced yet moderate accuracy. The three-pipe condition exhibited the best performance, with a Precision of 0.98 and Recall of 0.91, demonstrating strong model capability in identifying this class. The macro and weighted average F1-scores were 0.90 and 0.91 respectively, indicate consistent and robust performance across all classes. A detailed confusion matrix is shown in Fig. 4(c).

4. Analysis with deep learning model

In addition to traditional machine learning techniques, we employed CNN to analyze the data. CNNs are a class of deep learning models particularly suited for processing grid-like data structures, making them highly effective for image and spatial data analysis (LeCun *et al*. 2015, Chen *et al.* 2021). Because the resistance values in our dataset do not inherently form a grid-like or spatial structure, a transformation was required before feeding the data into the CNN model. We preprocessed the resistance values by reshaping them into a 2D grid format, enabling CNN to detect new features across the transformed data.

4.1 Data transform

In machine learning techniques, each resistance value is treated independently. We represented the resistance data in a structured format that captures the relationships between the electrodes. As illustrated in Fig. 5(a), the eight electrodes are considered nodes, with each resistance value representing the interaction or “relationship” between a grounded node and the other nodes. For each grounded node, there are seven resistance measurements



(a) $R_{i,j,}$ the relationship between node i (ground node) and node j

(b) Electrode Interaction Matrix

Fig. 5 Data transform

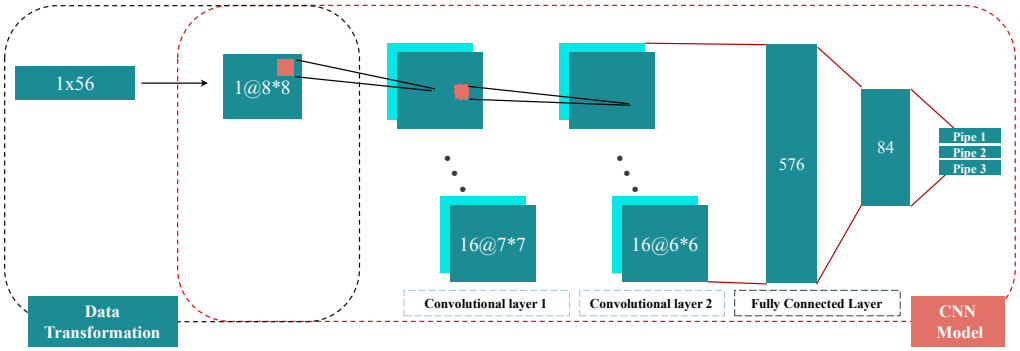


Fig. 6 CNN model: two convolutional layers and two fully connected layers

corresponding to its connections with the other nodes. Additionally, we included a self-relationship $R_{i,i}$, set to zero, representing no resistance within the same node. Consequently, each grounded node generated eight relationship values, representing its relationships with other nodes.

This structure enables the creation of a correlation matrix, as illustrated in Fig. 5(b). In this matrix, each element $R_{i,j}$ represents the resistance value between nodes i and j, with node i grounded. We organized the resistance values into a grid-like matrix format, where each row corresponds to a grounded node and its relationships with all other nodes. For instance, the first row represents the relationships when the first node is grounded, the second row corresponds to the second grounded node, and so forth.

Transforming the data into this grid-like structure enables the CNN model to leverage its strength in capturing local patterns and spatial relationships. CNNs excel at identifying complex interactions within structured data, making this approach well-suited for extracting meaningful features from the resistance values.

4.2 CNN model

The CNN model used to analyze the Electrode Interaction Matrix of the resistance data is relatively simple and designed to suit the dataset's characteristics and the study's objectives. As shown in Fig. 6, the model comprises two convolutional layers, followed by two fully connected layers. Each convolutional layer has 16 filters with a 2×2 kernel size, enabling the model to capture local patterns in the data. The two fully connected (dense) layers further process the extracted features. The final output layer contains three units, corresponding to the three classes in the classification task.

Table 6 CNN model summary

Layer(type)	Output Shape	Param #
Conv2d-1	*[−1, 16, 7, 7]*	*80*
Conv2d-2	*[−1, 16, 6, 6]*	*1,040*
Linear-3	*[−1, 84]*	*48,468*
Linear-4	*[−1, 3]*	*255*
Total params	*49,843*	
Trainable params	*49,843*	
Non-trainable params	*0*	

In total, the CNN model has 49,843 parameters, which are all trainable, as summarized in Table 6. The output shapes and parameter count for each layer reflect the model's compact structure, with the convolutional layers contributing to early feature extraction and the dense layers enabling classification.

This model architecture was deliberately kept simple for two main reasons. First, the dataset is relatively straightforward, representing interactions between

Table 7 Classification report of CNN model

Pipe Counts	Precision	Recall	F1-score	Support
1	0.83	0.96	0.89	82
2	0.97	0.77	0.86	78
3	0.96	0.99	0.96	95
Accuracy		0.91		255
Macro average	0.92	0.91	0.91	255
Weighted average	0.92	0.91	0.91	255

electrodes without highly complex patterns, and the classification task involves only three classes, reducing the need for a complex model. Second, the primary objective of this study is to demonstrate that the resistance data itself contains sufficient information and holds significant potential as a foundation for future research. By adopting a data-based approach, we show that reliable results can be achieved without increasing model complexity.

Using basic CNN architecture, we demonstrated that even a basic model can derive meaningful insights from the data, highlighting the value embedded in the transformed resistance matrix. This approach shows that with effective data transformation, significant information can be extracted for classification tasks, enabling accurate results without the need for complex or computationally intensive models.

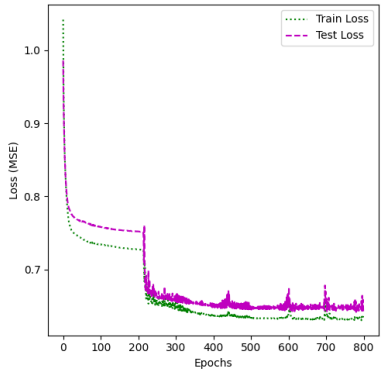
4.3 Results

The CNN model achieved an overall accuracy of 91%. For the one-pipe condition, Precision and Recall were 0.83 and 0.96, respectively, indicating strong performance in accurately identifying this class. The two-pipe condition showed a higher Precision of 0.97 but a slightly lower Recall of 0.77, suggesting a trade-off with some instances not being correctly identified. The three-pipe condition exhibited high Precision (0.96) and Recall (0.99), indicating highly accurate predictions with minimal missed instances. The macro and weighted average F1-scores were both 0.91, highlighting consistent and robust model performance across all classes. The classification report is provided in Table 7.

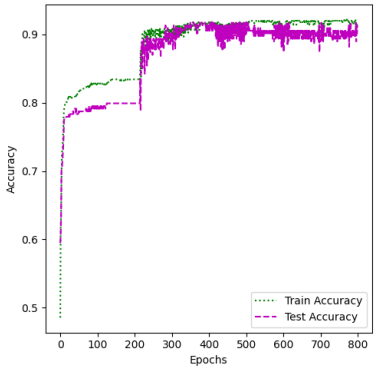
Fig. 7 shows the training and evaluation performance of the CNN model. In Fig. 7(a), the training and test loss curves exhibit a steady decline and eventual stabilization over the epochs, indicating effective learning and convergence of the model.Fig. 7(b) presents the accuracy curves for both the training and test sets, with accuracy increasing rapidly in the early epochs before stabilizing around 90%, demonstrating consistent performance throughout the training process. Fig. 7(c) displays the confusion matrix, which provides a detailed view of the model's classification performance across the three classes.

4.4 Comparison

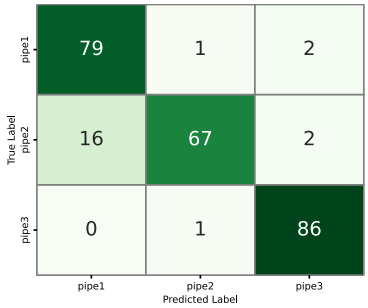
Fig. 8 presents a comparison of the performance metrics



(a) Loss Curve



(b) Accuracy Curve



(c) Confusion Matrix

Fig. 7 Loss curve and confusion matrix

for the models used in this study: SVM, XG-Boost, Random Forest, and CNN. Both CNN and Random Forest achieved the highest overall accuracy at 91%. However, CNN slightly outperformed Random Forest in Precision and F1-score, showing values of 0.92 and 0.91, respectively. XG-Boost achieved an accuracy of 87%, with balanced Precision, Recall, and F1-score values of 0.87 across all metrics. Although the SVM model was effective, it had the lowest accuracy at 85% and an F1-score of 0.85, indicating a slightly weaker performance than the other models.

Even though Random Forest is generally not considered a state-of-the-art technique, it exhibited competitive performance in this study. This can be attributed to its ensemble nature, which helps reduce variance and improve generalization. Additionally, the characteristics of the dataset, such as feature distributions and data complexity, may have favored tree-based ensemble methods. Proper hyperparameter tuning could have also contributed to its

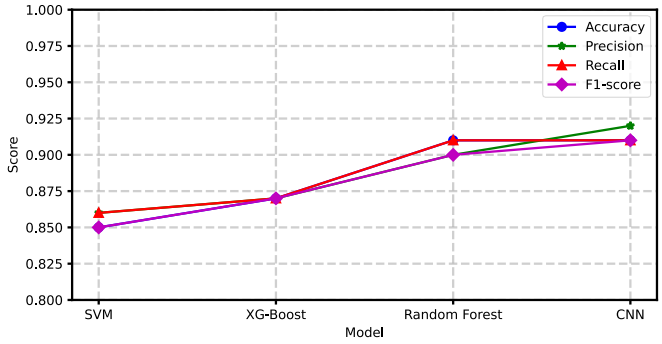


Fig. 8 Comparison of the four different models: SVM, XG-Boost, Random Forest, and CNN

strong performance. These findings suggest that, depending on the dataset and task, traditional ensemble methods can still offer reliable predictive accuracy.

Overall, the CNN model demonstrated the best performance in terms of Precision and F1-score, suggesting it is well-suited for capturing nuanced relationships in the data. This superior performance can be attributed to CNN's ability to automatically extract high-level features from raw input data, reducing the reliance on manual feature engineering. Additionally, CNNs excel in capturing spatial and hierarchical patterns within the data, which may have contributed to their strong predictive capability in this study.

5. Conclusions

This study explored the application of machine learning and deep learning models in analyzing electrical resistivity data for underground utility detection. The results demonstrate that the dataset contains sufficient information to characterize subsurface conditions, highlighting the potential of data driven approaches in geophysical surveys. Even with the current dataset, the models provided valuable insights into underground utilities, reinforcing their applicability in real-world scenarios.

By integrating numerical simulations with machine learning, this research establishes a predictive modeling framework that extends beyond simple count-based analysis. From an engineering perspective, the ability to estimate the number of buried utilities suggests a broader potential for spatial prediction, similar to Ground Penetrating Radar (GPR) applications. It is well known that raw GPR data cannot be directly used for analysis, as noise removal and preprocessing are essential steps. Similarly, processing field collected electrical resistivity data is a critical aspect that requires further research. Developing effective noise reduction techniques to refine field data and bring it closer to ideal simulated data remains a significant and challenging research topic. Although actual field data was not used, the study strongly suggests that its results have the potential to be applied to real-world problems, such as reducing excavation risks, preventing utility damage, and lowering infrastructure maintenance costs. Its societal impact lies in enhancing urban safety and enabling sustainable underground development.

This study confirms both the usability of electrical resistivity data and the applicability of machine learning techniques in underground utility detection. Moving forward, our research will focus on preprocessing field-collected data to ensure compatibility with the analytical methods established in this study. Additionally, further investigations can explore the prediction of pipeline burial locations and broader subsurface conditions, expanding the applicability of machine learning models in underground geophysical analysis.

Acknowledgments

This research was funded by Korea Electric Power Corporation, grant number R23SA01.

References

Abdelmawla, A., Ma, S., Yang, J.J. and Kim, S.S. (2023), "Subsurface anomaly detection utilizing synthetic GPR images and deep learning model", *Geomech. Eng.*, **33**(2), 203-209. https://doi.org/10.12989/gae.2023.33.2.203.

Ali, Z.A., Abduljabbar, Z.H., Taher, H.A., Sallow, A.B. and Almufti, S.M. (2023), "Exploring the power of eXtreme gradient boosting algorithm in machine learning: A review", *Academic J. Nawroz Univ.*, **12**(2), 320-334. https://doi.org/10.25007/ajnu.v12n2a1612.

Budiman, F. (2019), "SVM-RBF parameters testing optimization using cross validation and grid search to improve multiclass classification", *Научная визуализация*, **11**(1), 80-90. https://doi.org/10.26583/sv.11.1.07.

Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S. and Miao, Y. (2021), "Review of image classification algorithms based on convolutional neural networks", *Remote Sens.*, **13**(22), 4712. https://doi.org/10.3390/rs13224712.

Chhun, K.T. and Yune, C.Y. (2023), "Evaluation of strength characteristics of cement-stabilized soil using the electrical resistivity measurement", *Geomech. Eng.*, **33**(3), 261-269. https://doi.org/10.12989/gae.2023.33.3.261.

Herman, R. (2001), "An introduction to electrical resistivity in geophysics", *Am. J. Phys.*, **69**(9), 943-952. https://doi.org/10.1119/1.1378013.

Hong, C.H., Kim, J.S. and Chong, S.H. (2022), "Theoretical resistance in cylindrical electrodes with conical tip", *Geomech. Eng*., **30**(4), 337-343. https://doi.org/10.12989/gae.2022.30.4.337.

Jaufer, R.M., Ihamouten, A., Goyat, Y., Todkar, S.S., Guilbert, D., Assaf, A. and Dérobert, X. (2022), "A preliminary numerical study to compare the physical method and machine learning methods applied to GPR data for underground utility network characterization", *Remote Sens*., **14**(4), 1047. https://doi.org/10.3390/rs14041047.

Jeong, H.S., Abraham, D.M. and Lew, J.J. (2004), "Evaluation of an emerging market in subsurface utility engineering", *J. Constr. Eng. Management*, **130**(2), 225-234. https://doi.org/10.1061/(ASCE)0733-9364(2004)130:2(225).

Kaestner, C.A.A. (2013), "Support vector machines and kernel functions for text processing", Revista de Informática Teórica e Aplicada, 130-154. https://doi.org/10.22456/2175-2745.39702.

Kim, T.Y., Ryu, H.H., Kang, M., Choi, S. and Chong, S.H. (2024). "Effect of geometry of underground structure and electrode on electrical resistance measurement: A numerical study", *Geomech. Eng.*, **39**(1), 105-113. https://doi.org/10.12989/gae.2024.39.1.105.

LeCun, Y. and Bengio, Y. (1995), "Convolutional networks for images, speech, and time series", The handbook of brain theory and neural networks, **3361**(10), 1995.

Lee, J., Kim, K., Kang, M., Hong, E.S. and Choi, S. (2024), "Void detection for tunnel lining backfill using impact-echo method based on continuous wavelet transform and convolutional neural network", *Geomech. Eng*., **36(**1), 1-8. https://doi.org/10.12989/gae.2024.36.1.001.

Lew, J.J. (1996), "Subsurface utility engineering-an initial step in project development", *Proceedings of the Associated Schools of Construction*, Texas A&M University, Texas, USA.

Lew, J.J. (2000), Cost savings on highway projects utilizing subsurface utility engineering (No. FHWA-IF-00-014). United States. Department of Transportation. Federal Highway Administration.

Loke, M.H., Rucker, D.F., Chambers, J.E., Wilkinson, P.B. and Kuras, O. (2021), "Electrical resistivity surveys and data interpretation", In Encyclopedia of solid earth geophysics, 344-350. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-58631-7_46.

Olabode, O.P. and San, L.H. (2023), "Analysis of soil electrical resistivity and hydraulic conductivity relationship for characterisation of lithology inducing slope instability in residual soil", *Int. J. Geo-Eng.*, **14**(1), 7. https://doi.org/10.1186/s40703-023-00184-z.

Parmar, A., Katariya, R. and Patel, V. (2019), "A review on random forest: An ensemble classifier", *Proceedings of the International conference on intelligent data communication technologies and internet of things (ICICI) 2018*, 758-763. Springer International Publishing. https://doi.org/10.1007/978-3-030-03146-6_86.

Pisner, D.A. and Schnyer, D.M. (2020), Support vector machine. In Machine learning, 101-121. Academic Press. https://doi.org/10.1016/B978-0-12-815739-8.00006-7.

Reiter, D., Napoli, V., Cohen, J.P., Boone, S., Moseley, P., Alhasan, A. and Salerno, J. (2023), Availability, Feasibility, and Reliability of Available Nondestructive Evaluation (NDE) Technologies for Detecting and Locating Buried Utilities (No. FHWA-HRT-23-037). United States. Department of Transportation. Federal Highway Administration. Office of Infrastructure Research and Development.

Samadi, H., Hassanpour, J. and Rostami, J. (2024), "Assessment of shear strength of fine-grained and coarse-grained soil using actual EPB-TBM operating data", *Int. J. Geo-Eng*., **15**(1), 20. https://doi.org/10.1186/s40703-024-00220-6.

Samouëlian, A., Cousin, I., Tabbagh, A., Bruand, A. and Richard, G. (2005), "Electrical resistivity survey in soil science: A review", *Soil Tillage Res*., **83**(2), 173-193. https://doi.org/10.1016/j.still.2004.10.004.

Tariq, A., Abualshar, B., Deliktas, B., Song, C.R., Al-Nimri, B., Barret, B. and Glennie, N. (2024), "ANN-based evaluation system for erosion resistant highway shoulder rocks", *Int. J. Geo-Eng.*, **15**(1), 17. https://doi.org/10.1186/s40703-024-00216-2.

Todkar, S.S., Le Bastard, C., Baltazart, V., Ihamouten, A. and Dérobert, X. (2019), "Performance assessment of SVM-based classification techniques for the detection of artificial debondings within pavement structures from stepped-frequency A-scan radar data", *NDT & E Int.*, **107**, 102128. https://doi.org/10.1016/j.ndteint.2019.102128.

Tosti, F., Benedetto, A., Ciampoli, L. B., Lambot, S., Patriarca, C., and Slob, E.C. (2016), "GPR analysis of clayey soil behaviour in unsaturated conditions for pavement engineering and geoscience applications", *Near Surface Geophys.*, **14**(2), 127-144. https://doi.org/10.3997/1873-0604.2016011.

Uslu, B., Sinha, S.K. and County, H. (2014), Underground utility locating technology. IWA Publ. https://doi.org/10.2166/9781780405711.

Yehia, S., Abudayyeh, O., Nabulsi, S. and Abdelqader, I. (2007), "Detection of common defects in concrete bridge decks using nondestructive evaluation techniques", *J. Bridge Eng*., **12**(2), 215-225. https://doi.org/10.1061/(ASCE)1084-0702(2007)12:2(215).

Zembillas, N.M. (2003), "Subsurface utility engineering: a technology-driven process that results in increased safety, fewer claims, and lower costs", In New Pipeline Technologies, Security, and Safety, 1422-1428. https://doi.org/10.1061/40690(2003)163.