applied sciences

MDPI

Article

Solving Optimal Electric Vehicle Charger Deployment Problem

Seungmo Kim [1], Yeonho Jeong [2] and Jae-Won Nam [3,*]

[1] Department of Electrical and Computer Engineering, Georgia Southern University, Statesboro, GA 30460, USA; seungmokim@georgiasouthern.edu
[2] Department of Electrical, Computer and Biomedical Engineering, University of Rhode Island, Kingston, RI 02881, USA; yjeong@uri.edu
[3] Department of Electronic Engineering, Seoul National University of Science and Technology, Seoul 01811, Republic of Korea
* Correspondence: jaewon.nam@seoultech.ac.kr

Abstract: Electric vehicles (EVs) have already been acknowledged to be the most viable solution to the climate change that the entire globe has long been combating. Along the same line, it is a salient subject to expand the availability of EV charging infrastructure, which quintessentially necessitates the optimization of the charger's locations. This paper proposes to formulate the optimal EV charger location problem into a facility location problem (FLP). As an effort to find an efficient method to solve the well-known nonpolynomial deterministic (NP) hard problem, we present a comparative quantification among several representative solving techniques. This paper features two comprehensive case studies representing regions with an average and a high density of EVs. As such, this paper shows that the proposed framework can lead to successful location optimization with adequate refinement of solving techniques.

Keywords: electric vehicle (EV) charging; facility location problem (FLP); integer program; optimization solvers; machine learning

Citation: Kim, S.; Jeong, Y.; Nam, J.-W. Solving Optimal Electric Vehicle Charger Deployment Problem. *Appl. Sci.* **2024**, *14*, 5092. https://doi.org/10.3390/app14125092

Received: 16 May 2024
Revised: 3 June 2024
Accepted: 4 June 2024
Published: 12 June 2024

1. Introduction

It is widely agreed that electric vehicles (EVs) are a promising solution to relieve environmental issues. In fact, EV sales in the United States (U.S.) have increased yearly. Charging infrastructure is critical to the continued growth of EVs and its upstream industries. A lack of convenient and ubiquitous charging infrastructure is one of the key factors that impedes EV adoption [1]. The U.S. federal government moved swiftly to address this. An example is the EV Charging Action Plan [2,3] that provides USD 7.5 billion to develop 500,000 public chargers nationwide by the Year 2030. The private sector has also responded positively to the government's leading effort. As an example, Tesla announced a commitment to open thousands of its "Superchargers" to EVs made by other manufacturers [4]. Currently, Tesla provides 28,000 charging ports at Supercharger stations in the U.S., which have been accessible primarily to drivers of the company's own cars until now. Nonetheless, the reality still looks quite far-fetched. The U.S. government aims for the average availability of EV charging stations to be every 50 miles [5] (compared to gas stations existent every 3.5 miles [6]), while low-income residents are more dependent on automobiles and must travel further to access jobs and essential services [7]. To this end, this paper lays out a theoretical framework of optimal EV charger deployment.

We emphasize that our EV charger location optimization problem will very highly likely take a wide diversity of variables (including economic, societal, human behavioral, etc.) into account. As such, we set this paper to contribute to the following:

- Building a *comprehensive mathematical framework* accommodating the particular complexity,
- Demonstrating our numerical computational framework for solving the facility location problem (FLP) representing the optimal location;

- Laying out an extensive comparative study among the optimization *solving techniques* as an effort to find the most efficient solver;
- Applying the findings to two real-world *case studies* representing an average and high density of EVs.

2. Related Work

There have been studies on the subjects that are related to this paper's discussion. Table 1 summarizes the prior work that is related to this paper in an effort to highlight this paper's contribution.

Table 1. Comparison of this paper's contribution to prior work.

Literature	Contribution
FLP	Formulation into MINLP for various real-world problems
Distance Optimization	Stochastic analyses for location selection
Weight Assignment Techniques	Demand prediction by assigning weights to data
Machine Learning Techniques	More efficient weight assignment via priority prediction
Solving Techniques	Exact or heuristic approaches to solve NP-hard problems
This Paper	Comprehensive feasibility study encompassing the aforementioned numerical techniques

2.1. Problem Formulation Approaches

2.1.1. Facility Location Problem (FLP)

The latest research introduced a body of prior work that formulates the optimal deployment of EV chargers as a multi-objective optimization problem [8], considering various factors such as battery types [9] and distances to nearby energy sources [10].

We extended our investigation to the literature on facility location problem (FLP), which is analogous to our problem in the sense of having to determine optimal deployment plans including locations and expansion patterns [11]. It took our attention that the FLP is formulated as the mixed integer programming [12], wherein the constraints are reduced to continuous-variable linear equations. The literature goes on to generalize the formulation to the mixed integer linear programming [13] and the mixed-integer nonlinear programming (MINLP) [14].

2.1.2. Distance Optimization

This method aims to select locations that enable all potential users to receive services with the shortest distance or minimum cost [15]. Representative examples include the set cover, p-median, and p-center techniques. The set cover ensures that users receive similar levels of service, p-median aims to efficiently locate facilities with minimal movement required for service, and p-center minimizes the maximum distance between facilities and users. Distance optimization techniques allow for efficient location selection, but they require careful consideration of local information and feasibility of installation, distances between users and candidate locations, and the exact number of chargers to be installed. Consequentially, it is appropriate to use these techniques for the final location selection after considering all location analyses.

2.1.3. Weight Assignment Techniques

Another body of literature shed light on predicting demand in location selection by assigning weights to the data. Some examples of weight assignment techniques include the analytical hierarchy process (AHP) [16], standard analysis model-based location selection methods [17], and weighted average methods [15]. However, these methods often involve subjective elements, such as expert opinions and surveys, in the process of determining weights, which can be considered a disadvantage.

2.1.4. Machine Learning Techniques

Machine learning was found to provide more objective weight assignment by learning from numerical data and assigning weights to variables based on their importance in the learned model or using regression coefficients to determine priorities [18]. Furthermore, an abundant body of literature already justified the efficiency of machine learning in solving an optimization problem with an extremely large number of data points [19]. Therefore, as shall be detailed in Section 6, this paper adopts machine learning as a means to take advantage of its *scalability* in cases with a very large number of data points.

2.2. Solving Techniques

The key challenge in this research is an exhaustive enumeration that would quickly become computationally hopelessly expensive for the MINLP [20]. As such, solving an MINLP in polynomial time by using *exact* algorithms (e.g., cutting plane, branch and bound, etc.) can only be considered for small instances. For large instances, it is difficult to enumerate all the solutions due to the number of permutations that easily explodes to $n!$ [21], leaving fewer options such as the black-box solver [22].

Thus, for solving such large-instance NP-hard problems, *heuristic* approaches attract broader interests [21]. Considering our MINLP problem that is uniquely characterized by a large number of variables, we find it particularly suitable to adopt the *metaheuristic*, which is known as particularly efficient in solving combinatorial optimizations (which this research seeks to solve) by searching over a large set of feasible solutions with less computational effort, especially with incomplete or imperfect information or limited computation capacity [23]. Metaheuristics sample a subset of solutions that is otherwise too large to be completely enumerated or otherwise explored [24]. We performed a thorough investigation of the literature on metaheuristics, including the tabu search (TS) [25], random swaps [26], genetic algorithm [27], simulated annealing (SA) [28], ant colony [29], and memetic algorithms [30].

3. Problem Formulation

Here is how our problem is uniquely defined. The optimal locations of EV chargers will be found via solving an *FLP* [11]. We modify the traditional FLP such that it can further accommodate a wide diversity of factors (including economic, societal, human behavioral, etc.) depending on the context, to which the problem is applied.

3.1. Spatial Setup

We characterize the distribution of EV chargers as a *homogeneous Poisson point process (PPP)* over a finite two-dimensional space $\mathbb{R}^2$. As shall be detailed in Section 5.1, we deploy *EV chargers* and *demanding areas*, and find the connections from an EV charger to (a) demanding area(s). Predicated on the assumption of PPP, the locations of the chargers and demanding areas follow a *uniform distribution* on both X and Y axes.

This spatial setup forms the foundation for the key optimization problems that this paper targets to solve, viz., FLP for the optimal location.

3.2. Formulation to Capacitated FLP

Suppose there are n facilities and m customers. One wishes to choose (i) which of the n facilities to open, and (ii) which of the open facilities to use to supply the m customers, in order to satisfy some fixed demand at minimum cost. We propose to modify the canonical form of *capacitated FLP* [11] into an MINLP, which is given by

$$\min \overbrace{\sum_{i=1}^{n}\sum_{j=1}^{m} v_{ij}d_{ij}y_{ij} + \sum_{i=1}^{n} s_i x_i}^{\text{Classical capacitated FLP}} \\ \underbrace{+ \sum_{i=1}^{m}\sum_{j=1}^{n} \mathsf{E}_{ij}^{-1} + \sum_{i=1}^{m}\sum_{j=1}^{n} \Theta_{ij}(\theta_1, \cdots, \theta_N) + \cdots}_{\text{This work}} \tag{1}$$

$$\left.\begin{aligned} \text{s.t. } & \textstyle\sum_{i=1}^{n} y_{ij} \le 1 \\ & y_{ij} \ge 0, \quad \forall i, j = 1, \cdots, n \\ & x_i \in \{0, 1\}, \quad \forall i = 1, \cdots, n \end{aligned}\right\} \text{Classical capacitated FLP}$$

$$\left.\begin{aligned} & \mathsf{C}_{\min} x_i \le \textstyle\sum_{j=1}^{m} d_j y_{ij} \le \mathsf{C}_i x_i, \quad \forall i = 1, \cdots, n \\ & 0 \le \mathsf{E}_{ij}^{-1} \le 1, \quad \forall i, j = 1, \cdots, n \\ & \text{Any additional constraints on } \theta_1, \cdots, \theta_N \end{aligned}\right\} \text{This work}$$

We identify the parameters of our modified FLP as below:

- i and j are indexes for an EV charging facility and a demanding area (or, equivalently, a customer), respectively.
- v_{ij} gives the *variable cost* to obtain the electricity supplied to serve customer j.
- d_j gauges the *demand* from customer j.
- y_{ij} quantifies the fraction of the demand made by customer j and fulfilled by facility i.
- x_i indicates whether facility i *opens or not*.
- s_i denotes the *sunken cost* (also known as "fixed" cost) of opening a charging facility i.
- $\mathsf{E}_{i,j}$ defines the *equity* achieved at customer j via service from facility i.
- C_i and $\mathsf{C}_{\min}$ indicate the *capacity* of facility i and the required minimum capacity of any facility, respectively, both in the unit of kWh.

To elaborate on a few key variables, E_{ij} measures the equity by using the *Gini coefficient*, which ranges from 0 (i.e., complete equality) to 1 (i.e., complete inequality). As a measurement for the inequality of wealth or income [31], the Gini coefficient has also been used to measure how evenly the resource is allocated to the participants in a network [32]. We propose to use the coefficient as a gauge of *how many of the demands around a facility* are addressed.

As another means to pursue the equity, we propose a constraint with the minimum capacity for any charger, $\mathsf{C}_{\min}$. In fact, the State of Georgia has also adopted this idea in their EV charger deployment plan [33]. Moreover, it has been found that integrating a multitude of chargers at a single facility can contribute to lowering the sunken cost s_i [34].

By Θ_{ij}, we leave some room for the possible *addition of new variables* as the formulation evolves to reflect the reality more accurately. As an MINLP, the formulation given in Equation (1) can accommodate Θ_{ij} either in the linear or nonlinear form. Furthermore, each variable $\theta_1, \cdots, \theta_N$ can either be discrete or continuous.

3.3. Unique Challenges

We are aware that the MINLP has been a well-studied area over the last few decades. In particular, the size and complexity of IP problems being successfully solved have increased, mostly thanks to the continued development of relevant algorithms—e.g., branch and bound [35].

Nonetheless, we consider this research novel and significant, owing to the extreme complexity of the target problem. This complexity is mainly attributed to the following key reasons [36]:

- **C1:** *Large search spaces* for domain and other variables;
- **C2:** *Inexistence of polynomial-time* numerical solving. techniques

In regard to challenge **C1**, the focus of this research is to deal with a *a large number of variables*, which will be unavoidable to precisely quantify the equity E_{ij} reflecting all the demographic, geographic, and economic factors. The challenge here is the *dissimilarity* among the different data. As a remedy, we build on the literature of *coupled matrix and tensor factorization (CMTF)*, which jointly factorizes multiple datasets in the form of higher-order tensors and matrices by extracting a common latent structure from the shared mode [37].

Another focal point of this work is *addition of constraints* as a means to (i) cut off infeasible solutions [38] and/or (ii) linearization [39]. However, we will be especially sensitive in adding cuts to Equation (1)—which is already NP-hard. The reason is that, while it removes integer infeasibility, it can incur more constraints in each node of a branch-and-bound process, which can cause a higher delay in solving [20].

As a response to challenge **C2**, we focus on keeping our optimization framework in a *flexible* form, which is quintessential to accommodate more variables (some of which are even *unknown*!) as the framework evolves to reflect the real-world characteristics of our problem. In reality, many factors gauging the societal/economic/demographic equities can dynamically *change* both spatially and temporally as the society evolves [40]. This crucial necessity of flexibility significantly compounds the complexity to our formulation, to which we propose to understand the feasibility of countermeasure techniques including iterative methods [41], hierarchically structured space [42], sensitivity analysis [43], fuzzy decision analysis [44], and Monte-Carlo selection methods [45].

Particularly, in our problem, the factors forming the objective and constraints can dramatically change both spatially and temporally. That is, many factors gauging the societal/economic/demographic equities [40] are subject to form "temporality" and hence change over time as the society evolves [40]. This makes a strong case that the optimization problem and the solving method must be formulated into flexible forms so they can accommodate any addition/removal/change of parameters. It is another critical factor that makes our formulation more complicated.

4. Solving Techniques Development

4.1. Unique Challenges and Proposed Approaches

Recall that the FLP formulated as (1) is an NP-hard problem [46,47]. This entails critical challenges in *solving* the problem.

We have already identified the metaheuristic as the primary method to implement our problem (1), which would be the most suitable option considering the unique challenges of our problem **C1** and **C2**. In our problem, it is particularly important to find an approximate global optimum than to find a precise local optimum in a fixed amount of time, which makes a compelling case where the SA is preferable to exact algorithms (such as gradient descent, branch, and bound, etc.) [48].

Furthermore, algorithms with an aim to solve large-size instances of such combinatorial optimization problems apply *parallelism* for an expedition of both exact methods (e.g., the branch-and-bound algorithms) [49] as well as heuristics [20]. A representative example of the latter is parallelization of the objective function evaluation during a tabu search [50], for which graphical processing units (GPUs) [51] as well as the Compute Unified Device Architecture (CUDA) platform [52] have been used. As a means to efficient memory management [53], this research also seeks the feasibility of *cooperative heuristics* with a particular aim of improving the solution's speed and accuracy. The parallel instances can be executed via a global memory [54] or via distributed-memory systems [49]. It is noteworthy that the latter is efficient when the algorithms are independent among the parallel instances, and thus, no exchange of information across the memories is critical.

Many variants of the integer program (IP) are acknowledged to be NP-hard [55]. With this noted, this paper is devoted to developing *computational tools* for solving the proposed FLP problem, a well-acknowledged NP-hard IP problem [56].

4.2. Comparison among Solving Techniques

For a dedicated purpose of comparing a variety of numerical optimization solving techniques, we propose to modify Equation (1) into an abstract form of the *integer NLP*. Especially, we notice that the complexity is particularly induced by the nonlinearity. Thus, it is of critical importance to come up with a test function that suits to test the nonlinearity accurately.

We identify the *Rastrigin function* [57] that has long been known as a representative example, through which the performance comparison among the multitude of solving techniques can be clarified [58]. Considering the unique nature of containing a large number of variables defining the objective and constraints, we expand the Rastrigin function such that n is a sufficiently large number:

$$f_{\text{ras}}(\mathbf{x}) = An + \sum_{i=1}^{n} \left[x_i^2 - A\cos(2\pi x_i) \right] \tag{2}$$

where $A = 10$ and $x_i \in [-5.12, 5.12]$.

4.3. Alternative Techniques

We shed light on alternative approaches, considering that combinatorial optimization is notorious for being highly complex, and thus, one may end up having to find an "approximate" solution to the global optimal. Yet, we reiterate that SA, the method that we propose to adopt to solve the proposed optimization problem, is acknowledged for its ability to solve very complicated optimization problems even when exact methods fail [59]. Thus, it can still suffice what this research is looking for, in a practical sense.

Even despite this safe selection of method, in the event the optimal solution varies too widely, we plan to adopt statistical techniques that will help uphold the reliability of the proposed SA mechanism. An example technique is the principal component analysis (PCA) [60], which seeks to identify a certain set of factors that particularly dominate the solution of the QAP.

Parallel computing has also been attracting a considerable amount of research interest, thanks to its ability to distribute an extremely complex (and thus hopelessly challenging to solve) optimization problem into smaller instances and solve them instead. As such, it can be deemed an efficient strategy to utilize a parallel computing cluster. As an option for a further scale-up to a larger pool of servers, we suggest using the MATLAB (Ver. R2022b) Parallel Server [61].

Moreover, we highlight that this paper investigates a set of machine learning-based optimization techniques, which will be presented in Section 6.1. As shall be detailed in the section, such machine learning-based techniques turn out to be particularly efficient in finding optimal solutions in a very complicated problem, which often necessitates a numerical solving approach in lieu of a closed-form, theoretical one.

5. Case Study 1: Region with Average EV Density

Now, we lay out a case study through computational experiments. The first scenario that we investigate is a region with an average EV density. We identify the State of Georgia in the U.S. as an adequate geographic area for the case study, owing to its current density of EV charging infrastructure [62].

5.1. Case-Specific Refinement of Solving Method

Our first method to solve the FLP [11] is using an ILP solver. Figure 1 depicts an example mapping between 20 chargers and 40 demand areas, which are generated at random locations following a homogeneous PPP. The purple lines indicate which charger serves which demand areas. Note that the focus of this particular simulation is put showing how the problem is solved, rather than how to generate optimal facility locations. We used

MATLAB(Ver. R2022b)'s `intlinprog` solver, which finds the minimum of a constrained integer linear multivariate optimization.

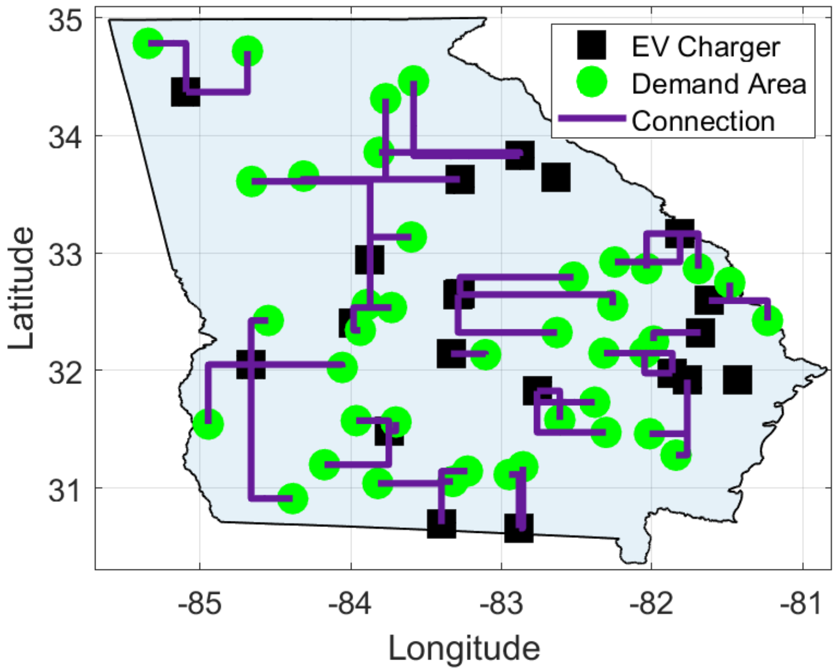


Figure 1. Example mapping of 20 EV chargers and 40 demand areas on the map of the State of Georgia.

As an effort to lay out a broader perspective on solving our problem, we make a comparison among a variety of numerical optimization methods. It is noteworthy that we refer to the Rastrigin function that has earlier been shown in Equation (2) with $n = 10$ as an effort to reflect the "many-variable" nature of our problem. Table 2 lays out the comparison. Note that the column labeled "Objective Value" shows the optimal value of the objective function, and columns x_1 through x_{10} indicate the values of x_i's in Equation (2) yielding the optimum.

Table 2. Comparison among several representative optimization solvers.

Solver	x_1	x_2	$\cdots$	x_{10}	Objective Value	Number of Iterations
Integer Linear Programming	4.4409×10^{-16}	4.4409×10^{-16}	$\cdots$	4.4409×10^{-16}	0	0
Pattern Search	0	0	$\cdots$	0	0	204
Genetic Algorithm	-0.062657	0.042974	$\cdots$	-0.041941	1.4801	3907
Particle Swarm	-7.2517×10^{-7}	2.5503×10^{-8}	$\cdots$	1.7757×10^{-6}	7.3×10^{-10}	4320
Simulated Annealing	6.4039×10^{-5}	-1.99	$\cdots$	0.00018799	3.9798	3008
Surrogate Optimization	0.99678	1.9937	$\cdots$	1.9832	8.9671	200

5.2. Results and Discussion

We expand the perspective of the proposed problem to taking *route minimization* into consideration. We propose to characterize route minimization as a traveling salesperson problem (TSP). Figures 2 and 3 demonstrate the EV chargers' locations on the map of the State of Georgia. The agent (i.e., salesperson) has to visit $n = 50$ points created on a map of the State of Georgia. The map is with shorter connections as iterations progress through the course of SA. Through the two figures, our SA algorithm (which has been shown as Algorithm 1) optimizes the route covering all the chargers and coming back to the starting point. The initial state in our algorithm is a connection to a completely random neighbor.

However, as the temperature is updated in each iteration, each node becomes able to connect to a closer neighbor. We notice here that our algorithm defines the temperature as the *distance* between two stops on the map.

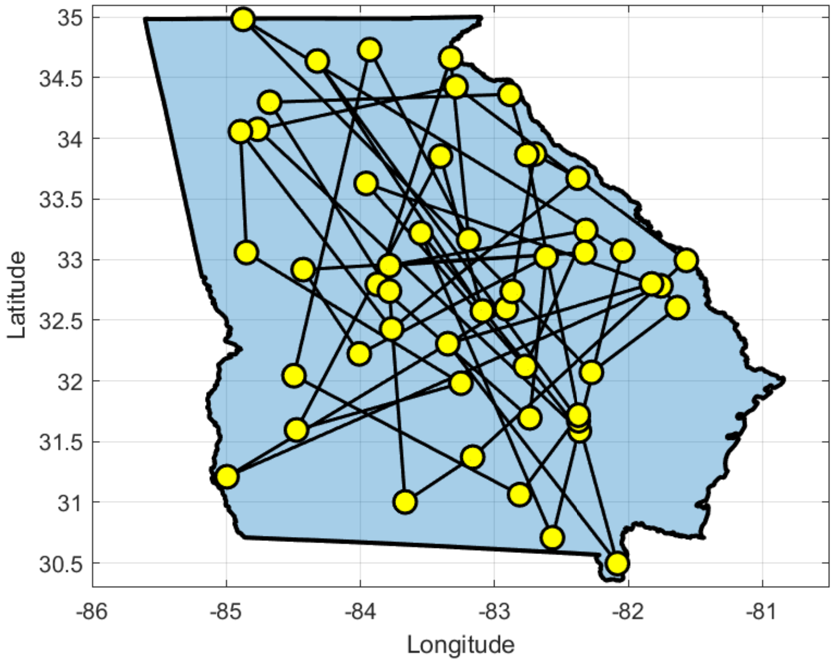


Figure 2. Solving a TSP by using SA (After 1 iteration).

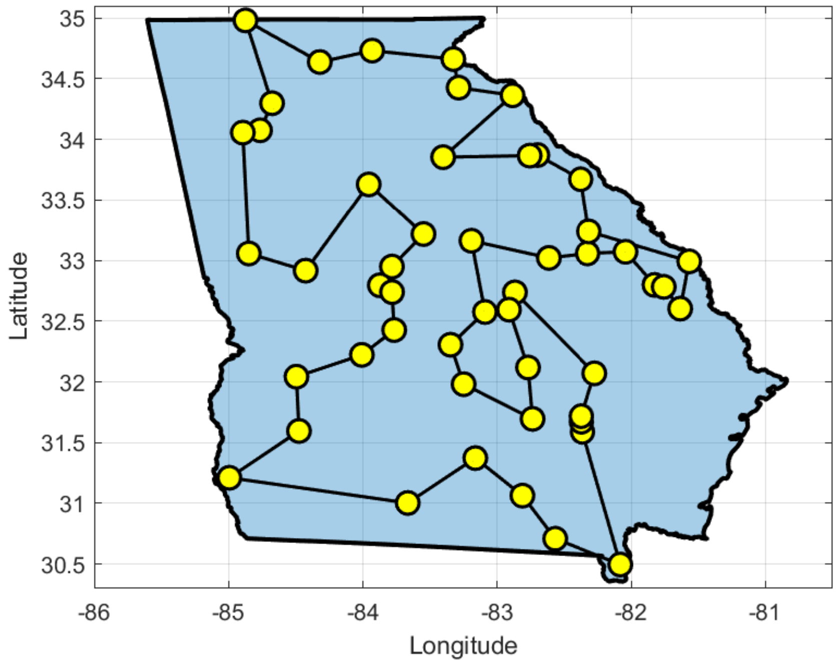


Figure 3. Solving a TSP by using SA (After 300 iterations).

Algorithm 1 SA implemented in this work

1: $s = s_0$
2: **for** $k \le k_{\max}$ **do**
3: $\quad \mathsf{T} \longleftarrow \text{temperature}((1-(k+1)/k_{\max}))$
4: $\quad s_{\text{new}} \longleftarrow \text{neighbor}(s)$
5: $\quad$ **if** $\mathsf{P}(\mathsf{E}(s), \mathsf{E}(s_{\text{new}}), \mathsf{T}) \ge \mathcal{U}(0,1)$ **then**
6: $\qquad s \longleftarrow s_{\text{new}}$
7: $\quad$ **end if**
8: **end for**

Algorithm 1 presents the pseudocode for the SA implementation in this simulation. The following parameters are used: T for the temperature, k for the index in the loop, $P(\cdot)$ for the acceptance probability, $E(\cdot)$ for the energy of a state, and U for the uniform random variable. The name of the algorithm "annealing" comes from the metallurgy process, through which a metal cools and freezes into a crystalline structure with minimum energy [59]. SA starts with an initial solution at a higher temperature T, where the changes are accepted with a higher probability P. Hence, the exploration capability of the algorithm is high, and the search space can be explored widely. As the algorithm continues to run, T decreases gradually, like the annealing process, and the acceptance probability of a non-successful move P decreases.

Figure 4 displays how the traveling distance converges as the iteration progresses via Algorithm 1.

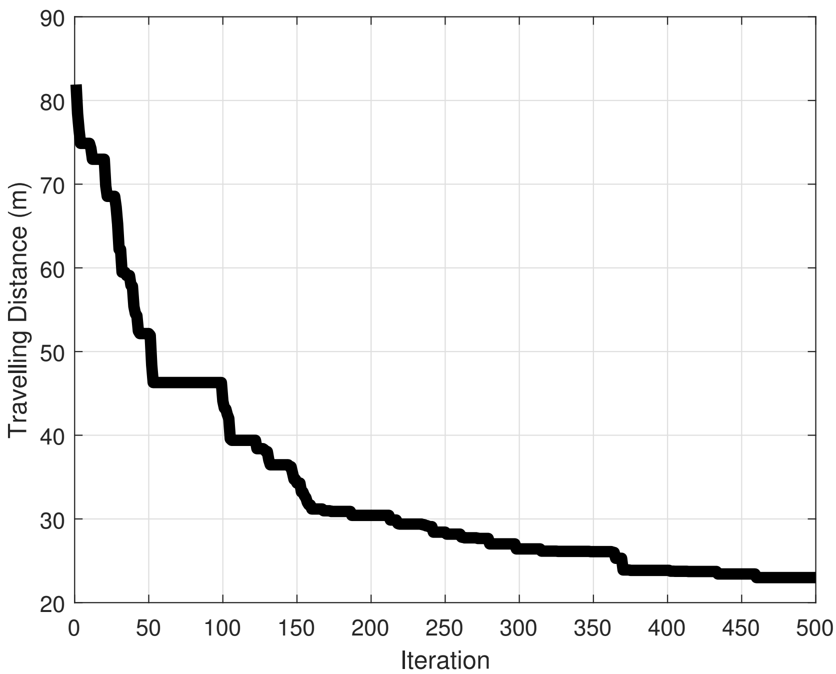


Figure 4. Convergence of total traveling distance.

6. Case Study 2: Region with High EV Density

In contrast to Section 5, this section is dedicated to presenting a case study of a region with a high EV density. We chose Jeju Island of South Korea, one of the places with a significantly higher level of EV penetration and thus the highest level of EV charger deployment [63,64].

Specifically, this section will elaborate our proposition to adopt a *machine learning*-aided approach as a means to solve the optimization problem with such an extremely large number of data points [19].

6.1. Case-Specific Refinement of Solving Method

In the proposition of the machine learning-based solving technique, several assumptions were established: (i) the currently installed charging stations are in the optimal location, and (ii) the selected locations of the current charging stations are influenced by various factors in the vicinity. Thus, in this study, the correlation between collected geographical or environmental data and the data of currently installed charging stations will be analyzed and compared to determine the optimal location selection.

6.1.1. Data Collection and Preprocessing

This case study features the use of QGIS [65], an open-source *geographic information* platform, in data preprocessing and visualization during the research.

We collect data on factors that are anticipated to influence the demand for charging stations and the data on currently installed EV charging stations. The collected data types include point, line, polygon, and grid data, totaling four types. For subsequent data merging, the line data will be divided into multiple point data with uniform values based on the point data as the reference. The polygon and grid data will be converted into point data with values at the center points of the grid based on a 250 m grid. Furthermore, considering that different types of buildings have varying impacts on charging station demand, we will group the building data based on the charging station utilization by building type data provided by the Korea Power Exchange [63]. EV charging stations can be broadly classified into fast-charging stations and regular charging stations, and this study focuses on public fast-charging stations for location selection.

Figure 5 illustrates the data collected in Jeju Island visualized in QGIS. The points in the figure have the same attribute values based on their colors, and it is assumed that a single EV charging station data point (shown in yellow) is influenced by nearby data points. Accordingly, the study will proceed with training based on this assumption.

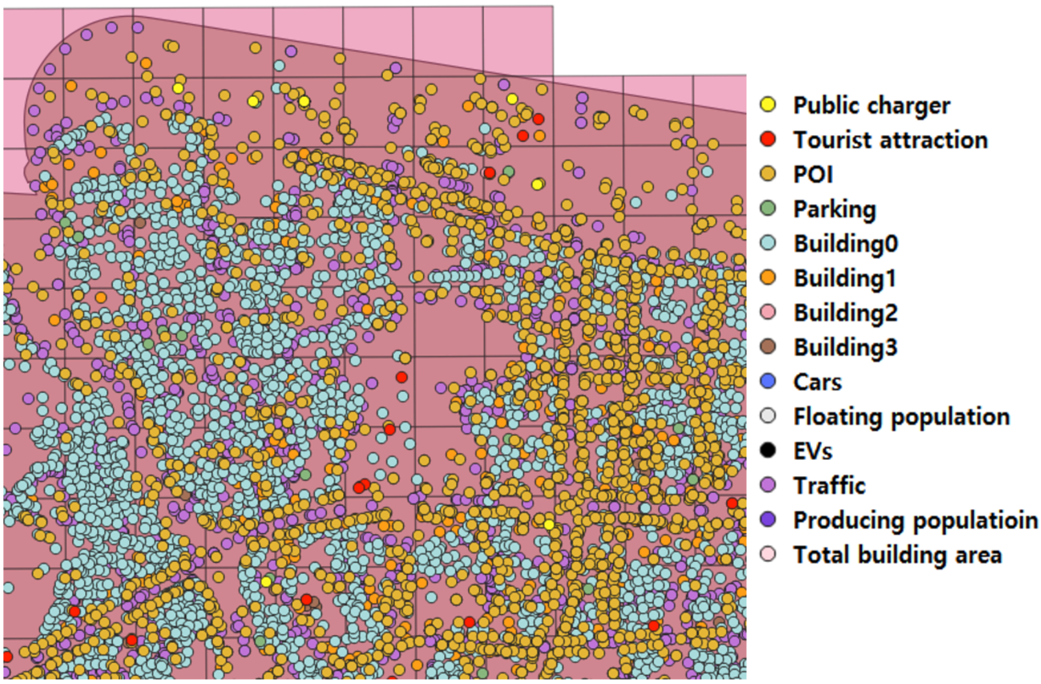


Figure 5. Partial visualization of the data in QGIS.

6.1.2. Data Integration

In [66], the authors conduct machine learning after setting a buffer, which represents the area where the reference data has a valid influence from its surrounding factors. A buffer indicates the extent of the category in which the reference data influences the surrounding data. Previous studies on site selection have used merged data based on grid cells, leading to the limitation of data having an influence only on the grid cell it belongs to. In contrast, the buffer-based approach considers the possibility of overlapping influences of data in the vicinity, making it a more reasonable method. Therefore, in this study, we proceed with the research by setting a buffer. The initial size of the buffer is determined based on the range of influence [64], which suggests that one charging station has an impact area of 4 km^2. Accordingly, we set the buffer radius to 1.13 km. Figure 6 illustrates the buffer with a radius of 1.13 km set around the charging station data (yellow dots).

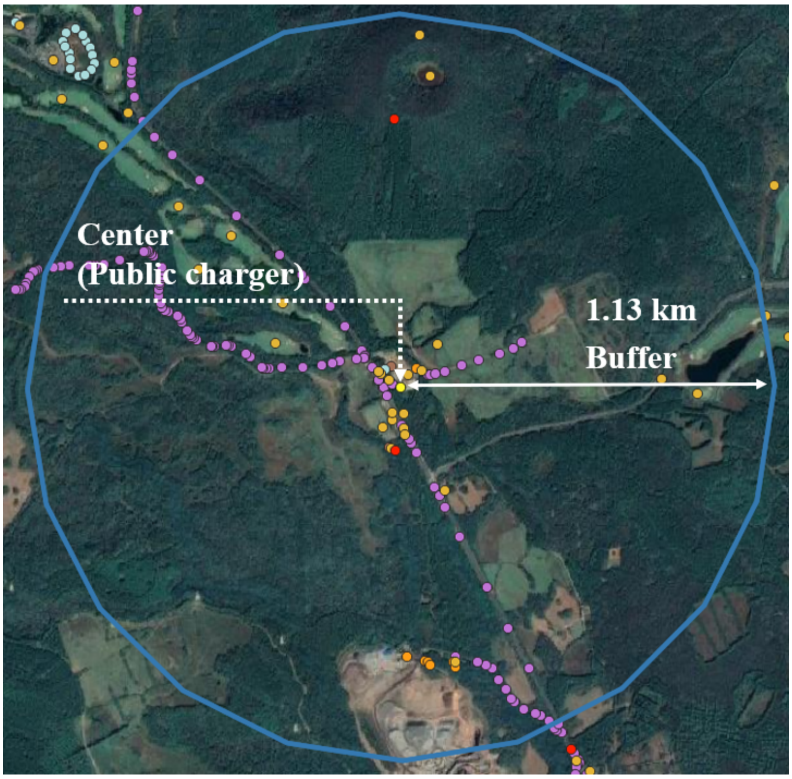


Figure 6. Visualization of the buffer in QGIS (See Figure 5 for the legend of each data dot).

Based on this buffer, we merge the data considering the data within the area has a valid influence on the charging station site selection. Additionally, to obtain sufficient training data, we add 10,000 random point data and merge them in the same manner as the charging station data.

6.1.3. Training Methods

As a means to train the machine learning models, we follow these steps:

1. *Data Preparation:* The collected and merged dataset undergoes preprocessing to ensure its suitability for training, including handling missing values, data normalization, and feature engineering.
2. *Model Selection:* The decision tree (DT) [67], support vector machine (SVM) [68], and random forest (RF) [69] models are considered as potential candidates for training due to their widespread usage in site selection problems.
3. *Training Process:* Each selected model is trained using the prepared dataset, which is divided into training and validation sets. Performance evaluation metrics such as accuracy, precision, recall, and F1 score are utilized to assess the model's performance during training.
4. *Model Evaluation:* After training, the models are evaluated using the validation set to assess their predictive capabilities. The evaluation metrics are used to compare the models' performance and identify the model with the highest accuracy or other desired performance metrics.

5. *Model Selection:* Based on the evaluation results, the model demonstrating the best performance is selected as the final machine learning model for the site selection task.

Following these steps ensures effective training of the machine learning models, enabling them to provide reliable predictions for optimal charging station locations.

To determine the model used for training, we compare the performance of commonly used models in site selection problems, namely, DT, SVM, and RF. Each model is trained with the data, and their performance is compared. (We justify the adoption of the three techniques as follows. This case study particularly targets to solve an optimal location problem in a region with a very high EV density. We found it adequate to rely on the methods that were proven to be computationally efficient in processing such an extremely large number of data points.)

For comparing the predicted optimal charging station locations using our proposed technique, we introduce the variable "consistency." Consistency measures the agreement between the predicted locations obtained through our approach and the current locations of installed charging stations in Jeju Island. It is normalized, ranging from 0 to 1. A consistency value of 1 indicates a perfect match, while 0 signifies no agreement between the predicted and actual locations. For instance, a consistency value of 0.8 means that our technique's predicted charging station locations match 80% of the current charging station distribution in Jeju Island, while the remaining 20% differ from our optimal predictions.

Table 3 presents the evaluation matrix for each model after training with the merged data. Based on these results, we select the Random Forest model, which exhibits the highest performance, to proceed with the site selection process. Using the selected model, we follow the steps outlined in Figure 7's flowchart to identify the conditions for site selection that yield high consistency.

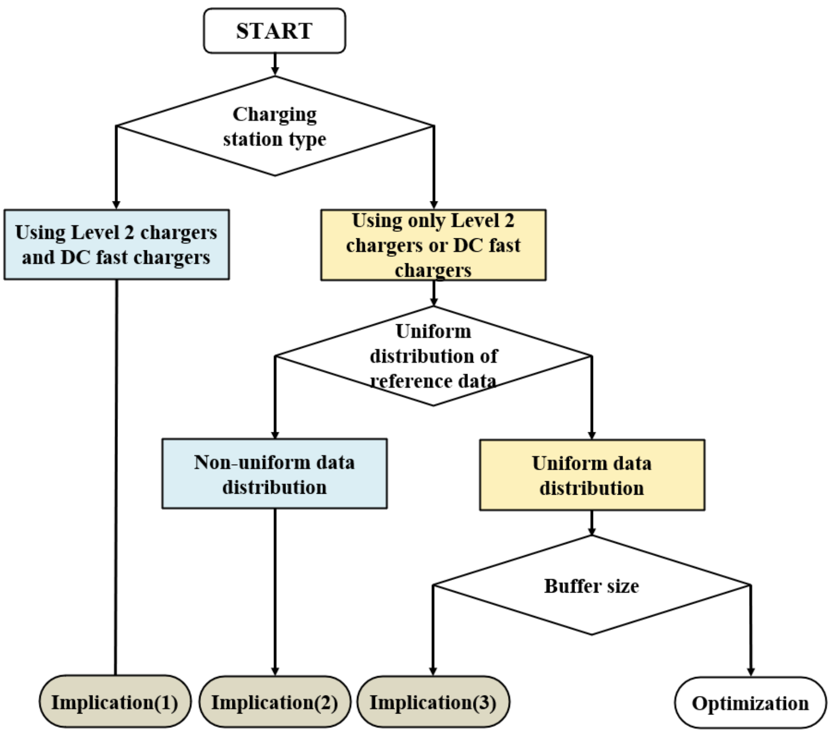


Figure 7. Algorithm of site selection.

Table 3. Evaluation matrix.

Model Name	Decision Tree	Support Vector Machine	Random Forest
Consistency	0.6583	0.4295	0.7580
Precision	0.6712	0.1845	0.7580
Recall	0.6583	0.4295	0.7580
F1-Score	0.6593	0.2581	0.7509

The conditions for each branching point are as follows:

(a) Charging Station Type: We examine the influence of the charging station type, which serves as the basis for training, on consistency. A comparison is made between training using only level-2 chargers or DC fast chargers and training with a mix of both.

(b) Uniform Distribution of Reference Data: The reference data initially set consists of charging station data which is not uniformly distributed across Jeju Island. To assess the impact of non-uniform data distribution on consistency, we compare the results of training based on 250 m grid points that are uniformly distributed as reference data with the results obtained from the original training.

(c) Buffer Size: We investigate the influence of the buffer size, which determines the scope of the reference data, and the resulting level of data duplication on consistency. To find the optimal buffer size, the previously used 1.13 km buffer size is set as the maximum value, and the buffer size is changed to measure the change in consistency. Figure 8 represents the buffers with a radius of 1.13 km (gray), 500 m (yellow), and 50 m (red).

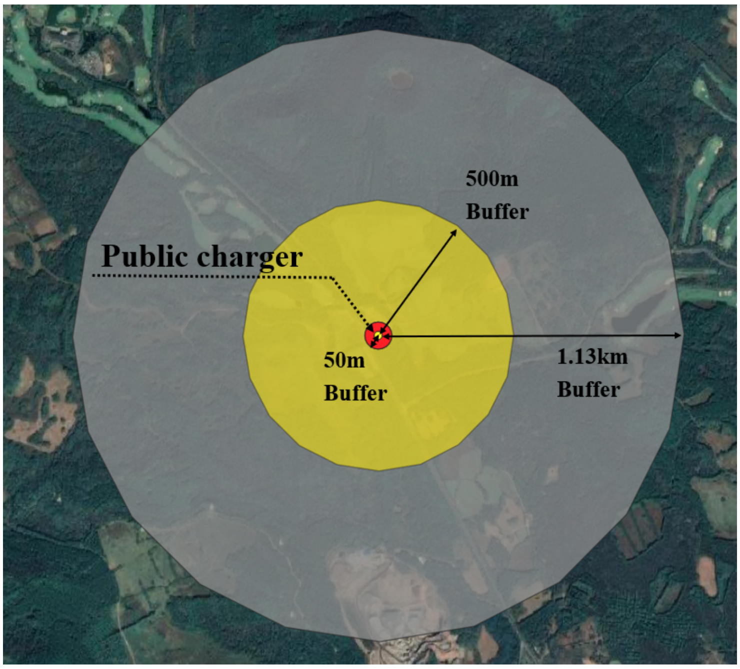


Figure 8. Size comparison of buffers.

6.2. Results and Discussions

Table 4 summarizes the configuration conditions and consistency results for each condition. The implications of these results are as follows:

1. Installation criteria differ between DC fast chargers and level-2 chargers. Significant differences in consistency are observed when training separately based on each charging station type or when training with both types together. Consequently, it can be concluded that chargers have been installed at locations that meet their respective criteria for both DC fast chargers and level-2 chargers.

2. Non-uniform distribution of reference data does not significantly affect training results. There is no significant difference in consistency between training based on non-uniformly distributed chargers and training based on grid points uniformly distributed at regular intervals. Thus, it can be concluded that the non-uniform distribution of data does not impact the training results.
3. Buffer size influences data consistency. Decreasing the buffer size results in increased consistency. The reason for the decrease in consistency at buffer sizes below 125 m is that the polygon data used for learning is 250 m × 250 m grid data, resulting in buffers that do not contain data from the 125 m radius buffer size. This problem can be solved by using smaller grid data than 250 m × 250 m grid data during data preprocessing. In conclusion, this result shows that larger buffer sizes increase data redundancy and affect consistency.

Table 4. Setting conditions and results of learning by condition.

	Setting Condition	Buffer Size	Consistency
Baseline	Public DC fast chargers and random point	1.13 km	0.7580
a. Charging Station Type	Public level-2 chargers and random point	1.13 km	0.7678
	Public DC fast chargers and public level-2 chargers and random point	1.13 km	0.6284
b. Uniform Distribution of Reference Data	Center point of grid data	1.13 km	0.7649
c. Buffer size	Public DC fast chargers and random point	700 m	0.8176
		600 m	0.8355
		500 m	0.8611
		400 m	0.8743
		300 m	0.9003
		200 m	0.9182
		150 m	0.9348
		125 m	0.9395
		100 m	0.9293
		50 m	0.8969

Based on the evaluation results, the Random Forest model demonstrates the highest consistency and is selected as the final machine learning model for the site selection task. Several conditions for site selection are analyzed, including the charging station type, uniform distribution of reference data, and buffer size. The results indicate that EV charging stations should be installed based on their respective criteria for both fast chargers and slow chargers. Furthermore, the non-uniform distribution of reference data does not significantly affect the training results, providing flexibility in data selection. Adjusting the buffer size influences data consistency, with smaller buffer sizes leading to increased consistency. By adjusting various conditions, we propose a method to achieve higher consistency in site selection through machine learning.

The final location selection is conducted based on the facts found from the setting conditions and results of the learning. First, since this study considers only public DC fast chargers, learning is conducted using only public DC fast chargers' data. In addition, the distribution of reference points for generating buffers was set as random points to give diversity in learning, considering that the plane data used for learning is uniformly distributed as 250 m × 250 m grid data. Moreover, despite the highest consistency obtained when the size of the buffer is 125 m, there are many areas that the buffer cannot cover. Therefore, a buffer of 500 m size, which has the highest match among the buffer sizes that include the whole of Jeju Island, is selected. Figure 9 shows the area containing 125 m and 500 m buffer.

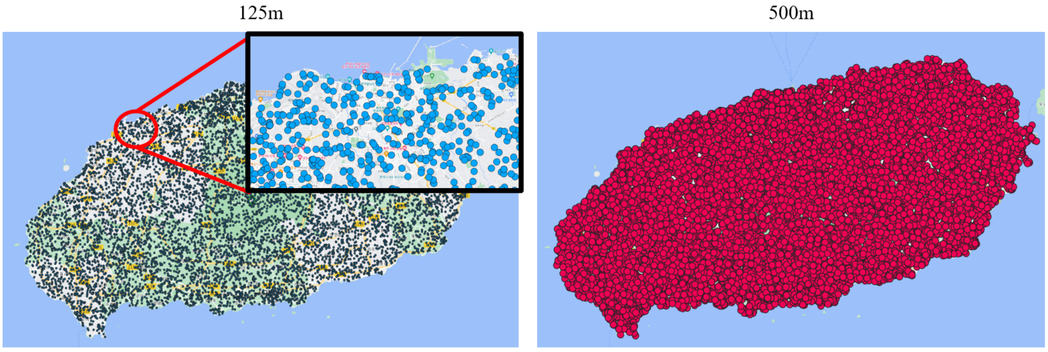


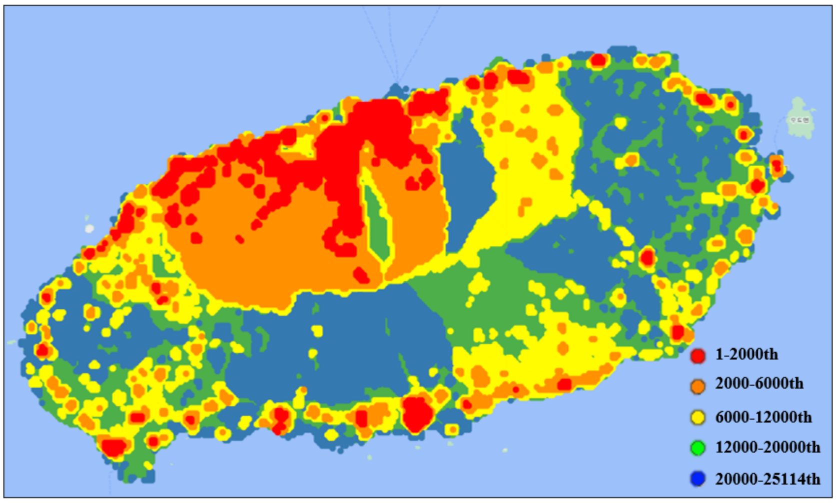
Figure 9. Area with 125 m (blue) and 500 m (red) buffers.

Machine learning was performed using the selected 500 m size buffer, and each importance is measured through the result of learning and used in the final location selection. Table 5 illustrates the importance of each calculated data.

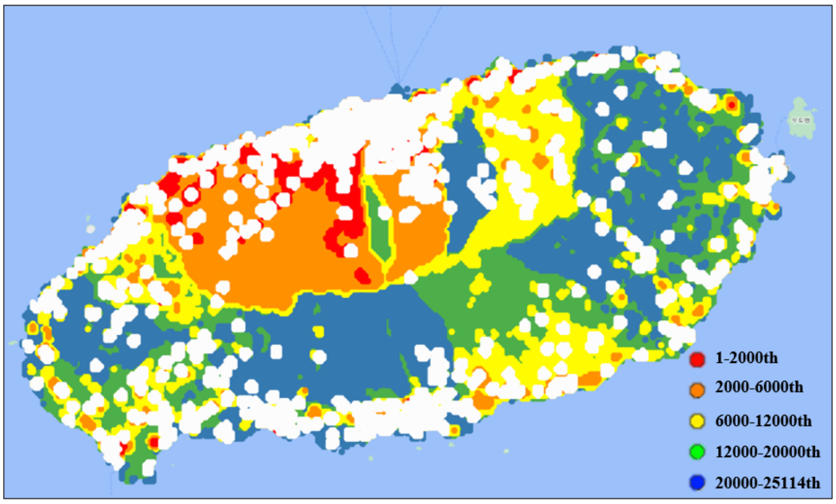
Table 5. Variable importance.

Data	Variable Importance [%]
POI	16.9264
Surface	12.9745
Building0	11.3435
Work_Population	9.4463
Building3	8.1137
Traffic	7.3983
Building1	6.768
Flow_Population	5.7931
Car	5.4492
EV_Car	5.3769
Parking	4.2025
Tour	3.563
Building2	2.6447

To make the final selection, the ranking is calculated by measuring the score based on variable importance and the data included in the buffer. The reference point for creating buffers to be used in the final location selection is the center point of the 250 m $\times$ 250 m grid to proceed with site selection for the entire Jeju Island. To calculate location ranking, normalize the data and measure the ranking based on the sum of the normalized data multiplied by the variable importance. If the values are the same, they are calculated in the same rank. Figure 10a shows the ranking of demand for public DC fast chargers across the entire Jeju Island based on the importance of each variable obtained from the machine learning results. Figure 10b presents the current DC fast charger locations added to the location rank distribution.



(**a**)



(**b**)

Figure 10. Location rank distribution. (**a**) Rank distribution, (**b**) EV charging stations installed in Jeju Island (white).

It shows that the current public DC fast charger installments and the high-ranking red and orange points generally match. Based on the above results, among the red points, which are the highest-ranking group, areas that do not overlap with the white points where public DC fast chargers are installed can be considered as target areas for additional installation of new public DC fast chargers.

7. Conclusions and Future Work

This paper has formulated an optimal EV charger location problem into the capacitated FLP. Then, it laid out an efficient method to solve the NP-hard problem. Via the first case study, we presented the feasibility of using an ILP solver to solve the proposed problem. The second case study presented the effectiveness of machine learning for selecting optimal locations for EV chargers and predicted the demand for EV chargers and compared it with the existing EV charger installations. As such, this study provides a foundation for conducting site analysis and selection research in the field of EV charger installations. In fact, Case Study 1 showed that our numerical solving method converged within 300 iterations even for a TSP, which is a well-known NP-hard problem. Moreover, Case Study 2 revealed that our machine learning-based optimization technique resulted in higher than 75% of consistency with the real deployment of EV chargers in the region.

The resultant response and management framework based on this paper's findings will provide a quantitative platform to balance the local, regional, and nationwide sustainability transition and resilience to fast EV expansion along the urban and rural continuum, thereby significantly contributing to balanced urban–rural EV charging infrastructure preparedness. As such, the findings will enlighten broader societies on a wide variety of practical issues that will likely be encountered during the broader deployment of EVs throughout the globe.

Considering this work's broad contribution, we will set the future work to expand onto various relevant avenues. One example can be revising the optimization problem reflecting the improvement in design of EVs and batteries. A longer-range coverage by a single EV may have potential to affect our problem formulation.

Author Contributions: Conceptualization, S.K.; methodology, S.K.; software, S.K.; validation, J.-W.N.; formal analysis, S.K. and Y.J.; investigation, J.-W.N., S.K. and Y.J.; resources, Y.J. and J.-W.N.; data curation, S.K. and J.-W.N.; writing—original draft preparation, S.K.; writing—review and editing, Y.J. and J.-W.N.; visualization, J.-W.N. and S.K.; supervision, S.K.; project administration, J.-W.N.; funding acquisition, J.-W.N. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Research Program funded by the SeoulTech (Seoul National University of Science and Technology: 2021-1135).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: Authors would like to thank Youlim Choi, Sunghyun Im, Hansoo Jung, for technical support (e.g., materials used for experiments).

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

References

1. Gersdorf, T.; Hensley, R.; Hertzke, P.; Schaufuss, P.; Tschiesner, A. *The Road Ahead for E-Mobility*; McKinsey: Chicago, IL, USA, 2020.
2. The White House. Fact sheet: The Biden-Harris electric vehicle charging action plan. In *Statements and Releases*; The White House: Washington, DC, USA, 2021.
3. United States Department of Transportation. *Biden-Harris Administration Announces $1.5 Billion Available through the 2023 RAISE Grant Program*; United States Department of Transportation: Washington, DC, USA, 2022.
4. Mitchell, R. U.S. government will pay Tesla to open its charger network to non-Tesla EVs. *Los Angeles Times*, 15 February 2023.
5. Newburger, E. *All 50 States Get Green Light to Build EV Charging Stations Covering 75,000 Miles of Highways*; CNBC: Englewood Cliffs, NJ, USA, 2022.
6. LNG2019. The Average Distance between Gas Stations in the United States. November 2022. Available online: https://www.linkedin.com/pulse/how-many-gas-stations-united-states-timothy-ohagan/ (accessed on 1 May 2024)
7. Bauer, G.; Hsu, C.; Lutsey, N. *When Might Lower-Income Drivers Benefit from Electric Vehicles? Quantifying the Economic Equity Implications of Electric Vehicle Adoption*; ICCT Working Paper; International Council on Clean Transportation: Washington, DC, USA, 2021.
8. Moghaddam, V.; Ahmad, I.; Habibi, D.; Masoum, M. Dispatch management of portable charging stations in electric vehicle networks. *ETransportation* **2021**, *8*, 100112. [CrossRef]
9. Onat, N.; Kucukvar, M.; Aboushaqrah, N.; Jabbar, R. How sustainable is electric mobility? A comprehensive sustainability assessment approach for the case of Qatar. *Appl. Energy* **2019**, *250*, 461–477. [CrossRef]
10. Qiao, B.; Liu, J. Multi-objective dynamic economic emission dispatch based on electric vehicles and wind power integrated system using differential evolution algorithm. *Renew. Energy* **2020**, *154*, 316–336. [CrossRef]
11. Jakob, K.; Pruzan, P. The simple plant location problem: Survey and synthesis. *Eur. J. Oper. Res.* **1983**, *12*, 41.
12. Ayazi, S.; Askarzadeh, A. Finding optimal path of feeder routing problem in power distribution network by an efficient and new methodology. *Int. Trans. Elect. Energy Syst.* **2021**, *31*, e13196. [CrossRef]
13. Floudas, C.; Lin, X. Mixed integer linear programming in process scheduling: Modeling, algorithms, and applications. *Ann. Operat. Res.* **2005**, *139*, 131–162. [CrossRef]
14. Shen, X.; Shahidehpour, M.; Zhu, S.; Han, Y.; Zheng, J. Multi-stage planning of active distribution networks considering the co-optimization of operation strategies. *IEEE Trans. Smart Grid* **2016**, *9*, 1425–1433. [CrossRef]
15. Kim, K.; Goo, Y. Optimizing fast EV charging infrastructure location with traffic flow data. *Innov. Stud.* **2020**, *15*, 61–91. [CrossRef]

16. Kim, J.; Lee, D.; Kim, S. A study to determine the optimized location for fast electric vehicle charging station considering charging demand in seoul. *J. Korea Inst. Intell. Transp. Syst.* **2022**, *21*. [CrossRef]
17. Yun, C. The case study of priority installation area of electric vehicle charging infrastructure using bigdata standard reference analysis model. *Proc. Symp. Korean Inst. Commun. Inform. Sci.* **2019**, 23–26.
18. Jo, J.; Kim, S. A Study on the optimal location selection for hydrogen refueling stations on a highway using machine learning. *J. Cadastre Land InformatiX* **2021**, *51*, 83–106.
19. Hernandez, A.; Perez, M.; Gupta, S.; Muntes-Mulero, V. Using machine learning to optimize parallelism in big data applications. *Future Gener. Comput. Syst.* **2018**, *86*, 1076–1092. [CrossRef]
20. Klotz, E.; Newman, A. Practical guidelines for solving difficult mixed integer linear programs. *Surv. Operat. Res. Manag. Sci.* **2013**, *18*, 18–32. [CrossRef]
21. Abdel-Basset, M.; Manogaran, G.; Rashad, H.; Zaied, A. A comprehensive review of quadratic assignment problem: Variants, hybrids and applications. *J. Ambient Intell. Human Comput.* **2018**. [CrossRef]
22. Zhang, H.; Beltran-Royo, C.; Ma, L. Solving the quadratic assignment problem by means of general purpose mixed integer linear programming solvers. *Ann. Operat. Res.* **2013**, *207*, 261–278. [CrossRef]
23. Blum, C.; Roli, A. Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Comput. Surv.* **2003**, *35*, 268–308. [CrossRef]
24. Bianchi, L.; Dorigo, M.; Gambardella, L.; Gutjahr, W. A survey on metaheuristics for stochastic combinatorial optimization. *Nat. Comput.* **2008**, *8*, 239–287. [CrossRef]
25. Drezner, Z. The extended concentric tabu for the quadratic assignment problem. *Eur. J. Operat. Res.* **2005**, *160*, 416–422. [CrossRef]
26. Misevicius, A. An implementation of the iterated tabu search algorithm for the quadratic assignment problem. *Operat. Res. Spectrum* **2012**, *34*, 665–690. [CrossRef]
27. Drezner, Z. Compounded genetic algorithms for the quadratic assignment problem. *Operat. Res. Lett.* **2005**, *33*, 475–480. [CrossRef]
28. Misevicius, A. An improved hybrid optimization algorithm for the quadratic assignment problem. *Math. Model. Anal.* **2004**, *9*, 149–168. [CrossRef]
29. Talbi, E.; Roux, O.; Fonlupt, C.; Robillard, D. Parallel ant colonies for the quadratic assignment problem. *Future Gener. Comput. Syst.* **2001**, *17*, 441–449. [CrossRef]
30. Benlic, U.; Hao, J. Memetic search for the quadratic assignment problem. *Expert Syst. Appl.* **2015**, *42*, 584–595. [CrossRef]
31. Gini, C. Variability and mutability. *J. Roy. Statist. Soc.* **1913**, *76*, 476–477.
32. Dai, L.; Jia, Y.; Liang, L.; Chang, Z. Metric and control of system fairness in heterogeneous networks. In Proceedings of the 2017 23rd Asia-Pacific Conference on Communications (APCC), Perth, Australia, 11–13 December 2017.
33. Williams, D. *Feds Approve Georgia DOT Plan for EV Charging Stations*; Georgia Public Broadcasting: Atlanta, GA, USA, 2022.
34. Nicholas, M. *Estimating Electric Vehicle Charging Infrastructure Costs across Major US Metropolitan Areas*; Working Paper 2019-14; ICCT: Washington, DC, USA, 2019.
35. Kleinert, T.; Labbe, M.; Ljubic, I.; Schmidt, M. A survey on mixed-integer programming techniques in bilevel optimization. *EURO J. Comput. Optim.* **2021**, *9*, 100007. [CrossRef]
36. Deng, X.; Lv, T. Power system planning with increasing variable renewable energy: A review of optimization models. *J. Clean. Prod.* **2020**, *246*, 118962. [CrossRef]
37. Schenker, C.; Cohen, J.; Acar, E. A flexible optimization framework for regularized matrix-tensor factorizations with linear couplings. *IEEE J. Sel. Topics Signal Process.* **2020**, *15*, 506–521. [CrossRef]
38. Burkard, R.; Fortuna, T.; Hurkens, C. Makespan minimization for chemical batch processes using non-uniform time grids. *Comput. Chem. Eng.* **2002**, *26*, 1321–1332. [CrossRef]
39. Asghari, M.; Fathollahi-Fard, A.; Al-e-hashem, S.M.; Dulebenets, M. Transformation and linearization techniques in optimization: A state-of-the-art survey. *Mathematics* **2022**, *10*, 283. [CrossRef]
40. Hardman, S.; Fleming, K.; Khare, E.; Ramadan, M. A perspective on equity in the transition to electric vehicle. *MIT Sci. Policy Rev.* **2021**, *2*, 46–54. [CrossRef]
41. Mandi, J.; Stuckey, P.; Guns, T. Smart predict-and-optimize for hard combinatorial optimization problems. *Proc. AAAI Conf. Artificial Intell.* **2020**, *34*, 1603–1610. [CrossRef]
42. Baluja, S.; Davies, S. *Using Optimal Dependency-Trees for Combinatorial Optimization: Learning the Structure of the Search Space*; Carnegie-Mellon University: Pittsburgh, PA, USA, 1997.
43. Hyde, K.; Maier, H.; Colby, C. A distance-based uncertainty analysis approach to multi-criteria decision analysis for water resource decision making. *J. Environ. Manag.* **2005**, *77*, 278–290. [CrossRef]
44. Estalaki, W.M.; Kerachian, R.; Nikoo, M. Developing water quality management policies for the Chitgar urban lake: Application of fuzzy social choice and evidential reasoning methods. *Environ. Earth Sci.* **2016**, *75*, 1–16. [CrossRef]
45. Madani, K.; Read, L.; Shalikarian, L. Voting under uncertainty: A stochastic framework for analyzing group decision-making problems. *Water Resour. Manag.* **2014**, *28*, 1839–1856. [CrossRef]
46. Fowler, R.; Paterson, M.; Tanimoto, S. Optimal packing and covering in the plane are NP-complete. *Inform. Proc. Lett.* **1981**, *12*, 133–137. [CrossRef]
47. Megiddo, N.; Tamir, A. On the complexity of locating linear facilities in the plane. *Operat. Res. Lett.* **1993**, *1*, 194–197. [CrossRef]
48. Kirkpatrick, S.; Gelatt, C., Jr.; Vecchi, M. Optimization by simulated annealing. *Science* **1983**, *220*, 671–680. [CrossRef]

49. Clausen, J.; Perregaard, M. Solving large quadratic assignment problems in parallel. *Comput. Opt. Appl.* **1997**, *8*, 111–127. [CrossRef]
50. Luong, T.; Melab, N.; Talbi, E. GPU computing for parallel local search metaheuristic algorithms. *IEEE Trans. Comput.* **2011**, *62*, 173–185. [CrossRef]
51. Sonuc, E.; Sen, B.; Bayir, S. A cooperative GPU-based parallel multistart simulated annealing algorithm for quadratic assignment problem. *Eng. Sci. Technol., Int. J.* **2018**, *21*, 843–849. [CrossRef]
52. Website of NVIDIA's CUDA Zone. Available online: https://developer.nvidia.com/cuda-zone (accessed on 10 October 2023).
53. Zhu, W.; Curry, J.; Marquez, A. SIMD tabu search for the quadratic assignment problem with graphics hardware acceleration. *Int. J. Prod. Res.* **2010**, *48*, 1035–1047. [CrossRef]
54. Abdelkafi, O.; Derbel, B.; Liefooghe, A. A parallel tabu search for the large-scale quadratic assignment problem. In Proceedings of the 2019 IEEE Congress on Evolutionary Computation (CEC), Wellington, New Zealand, 10–13 June 2019.
55. Ghosh, D.; Goldengorin, B.; Sierksma, G. Data correcting algorithms in combinatorial optimization. In *Handbook Combinatorial Optimization*; Springer: Boston, MA, USA, 2004; Volume 5, pp. 1–53.
56. Dantzig, G. *Linear Programming and Extensions*, 6th ed.; Princeton University Press: Princeton, NJ, USA, 1974; pp. 545–547.
57. Rastrigin, L. *Systems of Extremal Control*; Nauka: Moscow, Russia, 1974.
58. Katoch, S.; Chauhan, S.; Kumar, V. A review on genetic algorithm: Past, present, and future. *Multimed. Tools App.* **2021**, *80*, 8091–8126. [CrossRef] [PubMed]
59. Dupanloup, I.; Schneider, S.; Excoffier, L. A simulated annealing approach to define the genetic structure of populations. *Mol. Ecol.* **2002**, *11*, 2571–2581. [CrossRef] [PubMed]
60. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [CrossRef]
61. Website of MATLAB Parallel Server. Available online: https://www.mathworks.com/products/matlab-parallel-server.html (accessed on 1 February 2023).
62. Website of Georgia State Government. Available online: https://www.georgia.org/mobility-infrastructure#/analyze?show_map=true&region=US-GA (accessed on 12 May 2023).
63. Power Planning Department Demand Forecast Team. *Electric Vehicle and Charger Supply and Use Analysis*; Korea Power Exchange: Naju-si, Republic of Korea, 2021.
64. Kang, C.; Jeon, S. *A Study on Establishment of Proper Installation Criteria of Electric Vehicle Charging Station in Gyeonggi-Do*; Gyeonggi Research Institute: Suwon, Republic of Korea, 2017.
65. Website of QGIS 3.36. Available online: https://www.qgis.org/en/site/ (accessed on 12 October 2023).
66. Jeon, H.; Park, J.; Kim, H. A Study on the optimum location selection of electric vehicle fast charging station using GIS. *KSCE 2022 Conv.* **2022**, 558–562.
67. Charbuty, B.; Abdulazeez, A. Classification based on decision tree algorithm for machine learning. *J. Appl. Sci. Technol. Trends* **2021**, *2*, 20–28. [CrossRef]
68. Hearst, M.; Dumais, S.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Their App.* **1998**, *13*, 18–28. [CrossRef]
69. Breiman, L. Random forests. *Mach. Learn.* **2001**, 5–32. [CrossRef]