Article

Optimal Allocation for Electric Vehicle Charging Stations

Jiwon Lee ¹, Midam An ¹, Yongku Kim ² and Jung-In Seo ^{3,*}

¹ Department of Big Data Science, Halla University, Wonju 26404, Korea; orchid9913@naver.com (J.L.); anmid0522@gmail.com (M.A.)

² Department of Statistics, Kyungpook National University, Daegu 41566, Korea; kim.1252@knu.ac.kr

³ Department of Information Statistics, Andong National University, Andong 36729, Korea

* Correspondence: leehoo1928@gmail.com

Abstract: Currently, more than 30% of the fine dust generated in the Seoul metropolitan area is a pollutant emitted from automobiles such as diesel vehicles, and air pollution caused by this is becoming increasingly serious. In addition, the importance of electric vehicle distribution is increasing due to the strengthening of international environmental regulations on automobile exhaust gas and increasing the possibility of depletion of petroleum resources. This manuscript proposes a method for selecting an optimal electric vehicle charging station location in expanding charging facilities to activate electric vehicle distribution. For the sake of illustration, directions will be provided on how to select the best location for electric vehicle charging stations using data from Seoul, which has the best access. As the features, the number of living population and work force people and the number of guest facilities, which are determined to affect demand for quick charging, are considered. The missing values of the observed data are imputed based on the kriging technique from spatial correlation, and by segmenting the data through clustering, a representative technique of unsupervised learning, the characteristics of each cluster are examined and the characteristics of the clusters are identified. In addition, machine learning techniques such as the elastic net, random forest, support vector machine, and extreme gradient boosting are applied to examine the influence of the features used in predicting classes of data. In clustering analysis, the optimal number of clusters was determined to be 3 based on the heuristic and information-theoretic methods, and all the machine learning techniques considered showed that the number of work force population is the most important feature in predicting classes of data. All things considered from our results, it is reasonable to install quick electric vehicle charging stations in the places with the highest concentration of work force population and guest facility.

Keywords: electric vehicle charging; kriging; clustering; machine learning; feature importance

Citation: Lee, J.; An, M.; Kim, Y.; Seo, J.-I. Optimal Allocation for Electric Vehicle Charging Stations. *Energies* **2021**, *14*, 5781. <https://doi.org/10.3390/en14185781>

Academic Editor: Piedad Garrido Picazo

Received: 10 June 2021

Accepted: 14 July 2021

Published: 14 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

According to the 2019 World Air Quality Report, which surveyed the average concentration of fine particulate matter (PM_{2.5}) in 98 countries around the world, revealed that Korea's air pollution level is the 26th most serious in the world and that the average concentration of ultrafine dust in Korea is the highest among OECD countries. In particular, Seoul was found to be the worst in Korea at 24.8 $\mu\text{g}/\text{m}^3$. In Figure 1, panels A and B present average PM_{2.5} exposure by country during 2019 and PM_{2.5} annual mean ($\mu\text{g}/\text{m}^3$) in Seoul, respectively.

Currently, diesel cars are known to be the biggest cause of fine dust in the South Korean metropolitan area, and environmental pollution problems are becoming more serious as particulate matter (PM) caused by diesel cars is added to carbon monoxide (CO), hydrocarbons (HC), and nitrogen oxides (NO_x), which are harmful substances emitted from gasoline cars. If internal combustion locomotives are replaced with electric vehicles (EVs) that generate driving power by supplying electric energy from high-voltage batteries to electric motors, the air pollution problem caused by cars can be solved drastically because

they do not emit air pollutants and greenhouse gases. For this reason, on 18 December 2018, the Ministry of Trade, Industry and Energy announced that it has significantly raised its domestic supply target for EVs to 430,000 units by 2022 [1]. However, despite this supply policy, there are constant criticisms that they are reluctant to purchase eco-friendly cars due to insufficient charging stations. Therefore, in order to accelerate the introduction of EVs, it is important to expand the charging infrastructure, and it is necessary to install possible public chargers in places where the demand is high. That is, it is very important to increase charging stations but choose the optimal location for the efficient distribution of charging stations.

Figure 1. Global map of estimated PM2.5 exposure by country/region in 2019. Grey countries and regions indicate that these locations had insufficient PM2.5 data available for 2019 (left) and PM2.5 annual mean ($\mu\text{g}/\text{m}^3$) over 3 years (right).

Some authors studied allocation for EV charging stations to optimally expand the charging infrastructure. Frade et al. [2] studied on the optimal location of the EV charging stations for an area of Lisbon, the capital city of Portugal with a strong concentration of population and employment. Lam et al. [3] provided four solutions to tackle the EV charging station placement problem: iterative mixed-integer linear program (MILP), greedy approach, effective MILP, and chemical reaction optimization. Chung and Kwon [4] proposed a multi-period flow-refueling location model in order to find an optimal EV charging station location and applied the model to the real traffic flow data of the Korean Expressway network in 2011. Akbari et al. [5] introduced a function based on the distance constructed with Haversine Formula [6] to minimize the recharge cost of the EVs. Erbaş et al. [7] provided the Geographic Information System-based Fuzzy Multi-Criteria Decision Analysis (GIS-based FMDA) methodology to select an optimal EV charging station location based on the fuzzy Analytical Hierarchy Process (AHP) and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). Fredriksson et al. [8] provided an iterative approximation technique to solve the Route Node Coverage (RNC) problem that is used to identify the optimal location of charging stations for EVs in a road transport network. Dong et al. [9] provided a new approach to find optimal locations of new charging stations by bringing together a spatial point process statistical model of charging points locations and a mathematical maximum coverage location model.

However, in practice, not only the slow charger, but also quick charger introduction and expansion are also very important issues to activate EV distribution. For example, the slow charger takes an average of four to five hours to buffer one EV, which greatly reduces user convenience due to their lower turnover rate compared to gas stations. This makes users prefer the quick charger, especially when driving in the suburbs, where quick charging infrastructure is scarce, the demand for the quick charger is bound to increase.

In keeping with these issues, this paper focuses on optimal allocation methods for quick charging stations for building quick charging infrastructures unlike existing studies, and studies efficient customized strategies to find areas that are expected to be in high demand through clustering analysis and machine learning techniques. The optimal allocation is analyzed in three aspects. One is a spatial interpolation method to process the missing values of the considered features, and the other is a cluster analysis based on the features to create a customized strategy for each cluster and machine learning techniques to identify the important features that predict each cluster of data.

The rest of this paper is organized as follows. Section 2 develops a methodology that is applied to determine optimal locations for EV charging stations such as the spatial interpolation, clustering analysis, and machine learning methods producing feature importance. Section 3 provides analysis based on real data for illustrative purposes. Finally, conclusions are provided in Section 4.

2. Methodology

2.1. Spatial Interpolation

For interpolation of unmeasured points, we employ the kriging method, which is a method that can be applied to the analysis of physical phenomena or data that are distributed spatially or visually. It is known that this approach effectively predicts the value of an unmeasured point by grasping the distribution characteristics of given data and analyzing the correlation based on this.

The kriging analysis is modeled by defining the sample space of the variable of interest, collecting the data, and grasping the spatial interrelationship of the collected data, and there is a representative scale for this, namely, the variogram. Representatively, there are three types for the variogram: the spherical model, exponential model, and Gaussian model. The spherical model is generally the most used functional model, is expressed in the form of a third-order polynomial, and has the characteristic that the variogram value coincides with the sill value in the correlation distance. The exponential model is interpolated to approximate to the sill value and is widely used mainly in hydrological data. The Gaussian model is used when there is a strong correlation of data or when one wants to use a model with strong continuity.

In addition, the kriging technique includes simple kriging, the simplest method used under the assumption that the mean is known, ordinary kriging, which is widely used under the assumption that the mean is unknown but constant, and universal kriging, which is a method that takes into account the tendency of the mean. Among them, the ordinary kriging is the most widely used and it serves to estimate a value at a point of a region for which a variogram is known, using data in the neighborhood of the estimation location. The ordinary kriging is defined as follows.

Let $y(\mathbf{S}_i)$ are given observations at spatial coordinates $\mathbf{S}_i \in D, i = 1, \dots, n$, where n is the number of observed locations, and D is the geographical domain. Then, the ordinary kriging prediction [10] is given by

$$\hat{y}(\mathbf{S}_0) = \mathbf{w}(\mathbf{S}_0)^T \cdot \mathbf{y}, \quad (1)$$

where $\hat{y}(\mathbf{S}_0)$ is the kriging prediction at unmeasured points $\mathbf{S}_0$ and $\mathbf{w}(\mathbf{S}_0)^T$ is a vector of kriging weights $w_i(\mathbf{S}_0)$ for $i = 1, \dots, n$, which are obtained by minimizing the expected squared prediction error under an unbiasedness condition.

2.2. Clustering

Clustering analysis is a representative unsupervised learning method that groups n objects from multiple predictor $D = \{x_1, \dots, x_n\}$ values observed for each object into clusters with similar characteristics and analyzes the relationships between clusters by identifying the characteristics of the clusters formed. Information-theoretic methods as well as elbow and silhouette heuristic are considered to estimate the optimal number of clusters k .

In the heuristic method, the elbow method measures how homogeneity and heterogeneity within a cluster change for different values of k , and then finds k whose performance is weakened. To be more specific, the total within-cluster sum of squares (WSS) is calculated according to for each k , then k can be chosen at a point where brings less improvement to cluster homogeneity. However, it is sometimes ambiguous in that the point should be chosen subjectively. Alternatively, the silhouette method [11] can be used, which provides the optimal k is the one that maximizes the average silhouette of observations over a range

of possible values for k . In addition to these heuristic methods, 26 indices are considered to examine the relevant number of clusters.

For the information-theoretic method, the most commonly used indices include the Akaike's information criterion (AIC [12,13]), Bayesian information criterion (BIC [14]), and sample size-adjusted BIC (SABIC [15]). The AIC and BIC are defined as $-2 \log L + 2p$ and $-2 \log L + p \log n$, respectively, where L is the maximized likelihood, the last term in each is referred to as a penalty component that is penalized by the number of estimated parameters p to discourage overfitting of a model. In addition, the SABIC is defined as $-2 \log L + p \log((n+2)/24)$, which has been demonstrated in [16] as the most accurate (>90%) information criterion through simulation studies if there are at least 50 subjects per latent class.

For clustering, the latent profile analysis (LPA) that is a branch of the Gaussian finite mixture modeling is conducted with the estimated k in real data analysis, which is known that it is superior to a data-driven approach like k -means that uses distance algorithms.

2.3. Feature Importance

In general, the feature importance can be assessed in two ways: model-specific and model-agnostic. This subsection gives a brief description of a model-agnostic method [17] for assessment of the feature importance, which allows comparing the importance of features between models with different structures. The algorithm of [17] is as follows.

Let $\mathcal{L}(\hat{\mathbf{y}}, \mathbf{X}, \mathbf{y})$ be a non-negative loss function that quantifies goodness of fit of an estimated model based on the matrix of p features for all n observations $\mathbf{X} = (\mathbf{x}'_1, \dots, \mathbf{x}'_n)$, where $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ and $\hat{\mathbf{y}}$ is the predictions corresponding to the values of the target variable $\mathbf{y} = (y_1, \dots, y_n)'$ for the estimated model.

1. Compute the value of the loss function for the original data, $\mathcal{L}_0 = \mathcal{L}(\hat{\mathbf{y}}, \mathbf{X}, \mathbf{y})$.
2. For each feature, create the new data $\mathbf{X}^*$ by randomly permuting on n observations.
3. Compute predictions $\hat{\mathbf{y}}^*$ based on $\mathbf{X}^*$ from the fitted model.
4. Compute the value of the loss function for the new data, $\mathcal{L}^* = \mathcal{L}(\hat{\mathbf{y}}^*, \mathbf{X}^*, \mathbf{y})$.
5. Quantify the importance of each feature by calculating the difference $\mathcal{L}^* - \mathcal{L}_0$ or ratio $\mathcal{L}^* / \mathcal{L}_0$.

Note that the larger difference or higher ratio in Step 5 is greater reliance on features because it indicates that the performance of the model has worsened after permuting the values of the feature. That is, the larger the change in the performance, the more important is the feature.

In addition, as the permuting data in Step 2 is created including randomness, the results based on this may depend on the composition of the permuted values. Then, the associated uncertainty can be assessed by repeating the procedure N times.

3. Analysis

For the purposes of illustration for analyzing the optimal location of EV charging stations, we consider the number of living population and work force people and the number of guest facility in the Seoul Metropolitan Government, where public data are best opened.

The Seoul Metropolitan Government began to estimate the living population data in accordance with the recommendations of the United Nations (UN) to predict administrative needs such as urban planning and to utilize the basic data needed to improve the allocation of administrative services. Currently, the living population in the Seoul Metropolitan Government refers to all the people that exist in a specific area of Seoul at a specific time, which were estimated using public big data held by the Seoul Metropolitan Government and KT telecom data.

For analysis, data sets acquired from Seoul Open Data website on 18 June 2020 (<http://data.seoul.go.kr/>, accessed on 18 June 2020) are used. For simplicity, the following abbreviations are used:

#LP : The number of living population

#WP : The number of work force population

#GF : The number of guest facility

There are 24 autonomous “gu” districts in Seoul, divided into 424 administrative “dong” sub-units in Seoul. The 424 administrative “dong” is plotted in the left panel of Figure 2. Then, to understand the installation status of public quick chargers for EVs in Seoul, installation year and location of public quick EV charging stations using data on public quick chargers from the Public Data Portal (data.go.kr, accessed on 29 June 2020) are plotted in the right panel of Figure 2.

Figure 2. Four-hundred-and-twenty-four administrative “dong” of Seoul (left) and installation year and location of public quick EV charging stations (right).

The compactness of each feature is plotted in Figure 3, which indicates that features #WP and #GF have missing values in 68 and 82 administrative dong, respectively. To predict the values in such administrative dong with missing values, we employ the ordinary kriging prediction (1) that does not require estimating the mean of the random field from the data as mentioned before.

Figure 3. Compactness of #LP (left), #WP (middle), and #GF (right). The transparent regions are administrative dong with missing values and the red circle indicates location of EV charging stations.

First, to test whether there is spatial autocorrelation for the considered features, Moran’s I test [18] is conducted after log transform of the features. These results are reported in Table 1, which reveals that the features have significant positive spatial autocorrelation. For the variogram, an exponential model is chosen to fit the experimental semivariogram that provides the values of parameters for the spatial interpolation of kriging. Then, the missing values are replaced using the ordinary kriging prediction (1).

The 424 administrative dong with spatially interpolated data are grouped into similar types of groups based on features considered. To determine the optimal value of k , first the heuristic methods such as elbow method, silhouette method, and other 26 criteria are used. These results are plotted in Figure 4. In addition, the LPA that is a model-based technique is applied to detect the number of latent classes based on information-theoretic methods including AIC, BIC, and SABIC.

Table 1. Results for Moran's I test.

Var	Statistic	p-Value
#WP	0.300	0.000
#GF	0.129	0.000

For the LPA, four models with the following variance-covariance matrix are considered:

Model 1: Equal variances and covariances fixed to 0 structured as $\begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}$,

Model 2: Varying variances and covariances fixed to 0 structured as $\begin{pmatrix} \sigma_{1k}^2 & 0 & 0 \\ 0 & \sigma_{2k}^2 & 0 \\ 0 & 0 & \sigma_{3k}^2 \end{pmatrix}$,

Model 3: Equal variances and equal covariances structured as $\begin{pmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{12} & \sigma_3^2 \end{pmatrix}$,

Model 4: Varying variances and varying covariances structured as $\begin{pmatrix} \sigma_{1k}^2 & \sigma_{21k} & \sigma_{31k} \\ \sigma_{12k} & \sigma_{2k}^2 & \sigma_{23k} \\ \sigma_{13k} & \sigma_{12k} & \sigma_{3k}^2 \end{pmatrix}$.

The results for the information-theoretic methods are reported in Table 2.

Figure 4. Results for the elbow (**left**), silhouette (**middle**), and other 26 criteria (**right**) as heuristic methods.**Table 2.** Results for information-theoretic criteria.

Model	Class	AIC	BIC	SABIC
1	2	3055.4	3095.9	3064.2
1	3	2941.8	2998.5	2954.1
1	4	2900.8	2973.7	2916.5
2	2	3090.6	3143.3	3102.0
2	3	2886.8	2967.8	2904.4
2	4	2861.1	2970.4	2884.7
3	2	2883.5	2936.2	2894.9
3	3	2829.6	2898.4	2844.5
3	4	2795.5	2880.5	2813.9
4	2	2801.0	2877.9	2817.7
4	3	2765.5	2883.0	2790.9
4	4	2769.6	2927.6	2803.8

In Figure 4, the result for the elbow method suggests that $k = 3$ is the optimal number of clusters as it appears to be the bend in the knee (or elbow) and the silhouette method

shows that $k = 2$ is the optimal number of clusters maximizing the average silhouette values with three clusters coming in as second optimal number of clusters. In addition, in the case of the other 26 criteria, 11 of 26 criteria suggest $k = 3$ as the optimal number of clusters. Table 2 indicates that the version 3-class of Model 4 is the best model in terms of the AIC and SABIC, while the 2-class version of Model 4 is the best model in terms of the BIC. Considering the results and interpretability, $k = 3$ was chosen as the final number of clusters. Three clusters are denoted as “Class” 1, 2, and 3 for the unity of terms.

Then, we can naturally judge that the demand for quick charging stations will be high in places where the living population, working population, and guest facility are concentrated. To identify these characteristics of each Class, descriptive statistics and boxplots for 3 classes in each feature are reported in Figure 5, which shows the least frequency with 48 (11.3%) administrative dongs belonging to Class 2, the highest 235 (33.3%) administrative dongs belonging to Class 1 were found. In addition, the features #WP and #GF have the highest values for both mean and standard deviation (SD) in Class 2, while feature #LP do so in Class 3.

Figure 5. Descriptive statistics and boxplots for each feature by Class. The red triangle presents the mean value of each Class.

Here, to examine which features have a more significant impact on predicting classes of data, various machine learning techniques that can measure the feature importance based on the algorithm provided in Section 2.3 are applied to observed data with Class determined through clustering analysis as the target variable. As the machine learning techniques, the elastic net (EN), random forest (RF), support vector machine (SVM), and extreme gradient boosting (XGB) are considered. To avoid overfitting and to tune the model's hyperparameters during training, 5-fold cross-validation is done by splitting the original training data into 5 partitions. To find the optimal hyperparameters, the accuracy is used as the performance measure, which is computed by averaging 5 accuracy values from repeated 5 times. The accuracy from the models with the optimal hyperparameters is reported in Table 3, indicating that the SVM has the best performance, and that there is no significant difference in performance between the remaining models. For the RF and XGB, the out-of-bag (OOB) error in addition to the cross-validation is applied to assess performance of the model. For the loss function (that is, model performance measure), the cross-entropy is considered as it is the multinomial classification solution. To take into account the uncertainty, the mean values over 50 permutations are computed. The results for the feature importance from the models are present in Figure 6. In Figure 6, the length of the bar corresponds to the mean value of difference $\mathcal{L}^* - \mathcal{L}_0$.

Table 3. The accuracy from the machine learning models with the optimal hyperparameters.

	EN	RF	SVM	XGB
Accuracy	0.907	0.914	0.967	0.910

Figure 6 indicates that %WP is the most important feature in all considered models, followed by %GF and %LP except for the RF. From the result, we can consider Class 2 that is the place where working population and guest facility are concentrated the most as the optimal locations for EV charging stations.

Figure 6. Means over 50 permutations of variable-importance measures for the machine learning models with the cross entropy as the loss function.

4. Conclusions

This paper proposes a methodology for optimal quick charging station allocation to build an efficient EV quick charging station infrastructure to expand the distribution of EVs, an alternative to diesel vehicles, which are the main cause of environmental pollution.

The proposed method is implemented in three aspects: The first step is a spatial interpolation method through a kriging technique using spatial correlation for interpolation for unmeasured points. Then, by performing clustering analysis such as the LPA for an efficient customized strategy, we can find areas of high demand from users. Finally, machine learning techniques are applied to determine which features play an important role in predicting clusters of data. To examine the validity of the proposed method, it was applied to the Seoul Metropolitan Government, which has the easiest public data access.

Through this approach, it is possible to select the areas that correspond to priority where quick charging stations for EVs are needed by examining important features in areas with high demand, which makes efficient resource allocation possible. In addition, it can reduce uncertainty about the demand by identifying which features play a more important role in areas where customer demand is expected to be high.

As prospective research, a study is underway on the actual required number of chargers at each charging station based on the estimated charging demand of the selected candidate areas and the charging amount prediction considering the time and day of the week. In addition, although this study was proposed for the study of selecting the

optimal location for EV charging stations, it can be applied to similar studies for the optimal location.

Author Contributions: J.-I.S. conceived and designed the research; J.-I.S. and J.L. analyzed the data and interpreted the results; J.L., M.A., Y.K. and J.-I.S. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Education) (No. 2019R111A3A01062838).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in the analysis can be obtained from <http://data.seoul.go.kr/>.

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

References

1. The Ministry of Trade, Industry and Energy. Available online: <http://www.me.go.kr/home/web/board/read.do?boardMasterId=1&boardId=930050&menuId=286> (accessed on 8 August 2020).
2. Frade, I.; Ribeiro, A.; Gonçalves, G.; Antunes, A.P. Optimal location of charging stations for electric vehicles in a neighborhood in Lisbon, Portugal. *Transp. Res. Rec.* **2011**, 2252, 91–98. [\[CrossRef\]](#)
3. Lam, A.Y.S.; Leung, Y.W.; Chu, X. Electric vehicle charging station placement: Formulation, complexity, and solutions. *IEEE Trans. Smart Grid* **2014**, 5, 2846–2856. [\[CrossRef\]](#)
4. Chung, S.H.; Kwon, C. Multi-period planning for electric car charging station locations: A case of Korean Expressways. *Eur. J. Oper. Res.* **2015**, 242, 677–687. [\[CrossRef\]](#)
5. Akbari, M.; Brenna, M.; Longo, M. Optimal locating of electric vehicle charging stations by application of genetic algorithm. *Sustainability* **2018**, 10, 1076. [\[CrossRef\]](#)
6. Jaramillo, J.H.; Bhadury, J.; Batta, R. On the use of genetic algorithms to solve location problems. *Comput. Oper. Res.* **2002**, 29, 761–779. [\[CrossRef\]](#)
7. Erbaş, M.; Kabak, M.; Özceylan, E.; Çetinkaya, C. Optimal siting of electric vehicle charging stations: A GIS-based fuzzy multi-criteria decision analysis. *Energy* **2018**, 163, 1017–1031. [\[CrossRef\]](#)
8. Fredriksson, H.; Dahl, M.; Holmgren, J. Optimal placement of charging stations for electric vehicles in large-scale transportation networks. *Procedia Comput. Sci.* **2019**, 160, 77–84. [\[CrossRef\]](#)
9. Dong, G.; Ma, J.; Wei, R.; Haycox, J. Electric vehicle charging point placement optimisation by exploiting spatial statistics and maximal coverage location models. *Transp. Res. Part D* **2019**, 67, 77–88. [\[CrossRef\]](#)
10. Goovaerts, P. *Geostatistics for Natural Resources Evaluation*; Oxford University Press: New York, NY, USA, 1997.
11. Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*; Wiley: New York, NY, USA, 1990.
12. Akaike, H. Information Theory and An Extension of the Maximum Likelihood Principle. In *Second International Symposium on Information Theory*; Petrov, B.N., Caski, F., Eds.; Akademiai Kiado: Budapest, Hungary, 1973; pp. 267–281.
13. Akaike, H. Factor analysis and AIC. *Psychometrika* **1987**, 52, 317–332. [\[CrossRef\]](#)
14. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, 6, 461–464. [\[CrossRef\]](#)
15. Sclove, S.L. Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika* **1987**, 52, 333–343. [\[CrossRef\]](#)
16. Yang, C.C. Evaluating latent class analysis models in qualitative phenotype identification. *Comput. Stat. Data Anal.* **2006**, 50, 1090–1104. [\[CrossRef\]](#)
17. Fisher, A.; Rudin, C.; Dominici, F. All models are wrong, but many are useful: Learning a variable’s importance by studying an entire class of prediction models simultaneously. *J. Mach. Learn. Res.* **2019**, 20, 1–81.
18. Moran, P.A.P. Notes on continuous stochastic phenomena. *Biometrika* **1950**, 37, 17–23. [\[CrossRef\]](#) [\[PubMed\]](#)