WILEY | Hindawi

Research Article

Analysis of Travel Mode Choice in Seoul Using an Interpretable Machine Learning Approach

Eui-Jin Kim (ID)

Institute of Construction and Environmental Engineering, Seoul National University, Seoul 08826, Republic of Korea

Correspondence should be addressed to Eui-Jin Kim; kyjcwal@snu.ac.kr

Understanding choice behavior regarding travel mode is essential in forecasting travel demand. Machine learning (ML) approaches have been proposed to model mode choice behavior, and their usefulness for predicting performance has been reported. However, due to the black-box nature of ML, it is difficult to determine a suitable explanation for the relationship between the input and output variables. This paper proposes an interpretable ML approach to improve the interpretability (i.e., the degree of understanding the cause of decisions) of ML concerning travel mode choice modeling. This approach applied to national household travel survey data in Seoul. First, extreme gradient boosting (XGB) was applied to travel mode choice modeling, and the XGB outperformed the other ML models. Variable importance, variable interaction, and accumulated local effects (ALE) were measured to interpret the prediction of the best-performing XGB. The results of variable importance and interaction indicated that the correlated trip- and tour-related variables significantly influence predicting travel mode choice by the main and cross effects between them. Age and number of trips on tour were also shown to be an important variable in choosing travel mode. ALE measured the main effect of variables that have a nonlinear relation to choice probability, which cannot be observed in the conventional multinomial logit model. This information can provide interesting behavioral insights on urban mobility.

1. Introduction

The recent emergence of new travel modes such as ride-sourcing, ride-hailing, and autonomous vehicles and the evolution of new mobility services such as mobility as a service and mobility on demand (known as MaaS and MoD, respectively) is changing travel behavior significantly [1]. These emerging technologies present new sources of big data for understanding travel behavior and system performance [2]. New methods that leverage this big data are needed to analyze travel behavior changes and predict travel mode choices. The multinomial logit (MNL) model has dominated travel mode choice analysis due to its simplicity and readability. The simple MNL model and its variants have been applied to consider various effects in the context of travel mode choice based on the expert-designed model assumptions. Linear relationships in parameters of the simple MNL model can be intuitively interpreted as weights of the variables. Even nonlinear relationships in parameters such as

willingness-to-pay for reduced travel time variability can be captured by combining the conventional utility functional form with a probability weighting function [3]. However, this approach requires prior assumptions for the functional form of the weighting function. The MNL can capture the interaction effects between correlated variables by adding appropriate interaction parameters that are based on empirical or experimental knowledge [4], but considering all of the interactions becomes impossible as the number of variables increases. Although the simple MNL model assumes the independence of irrelevant alternatives (IIA) causing misleading predictions, the correlations between travel modes have been addressed by the advanced structure of the MNL model such as the nested logit and mixed logit model [5]. However, it is very difficult to design an appropriate model structure of the MNL model that effectively captures a high degree of complexity in a dataset [6]. In summary, the existing MNL and its variants can take into account the various effects in the mode choice situations;

however, they rely on the model assumptions that should be determined by the subjective judgment of the researcher, and these assumptions affect the parameter estimates and the prediction performance.

Machine learning (ML) approaches are promising alternatives to the MNL-based model for modeling travel mode choice. It can represent complex relationships between mode choices and input variables in a data-driven manner rather than making strict assumptions about the data [7]. Many previous studies have reported the use of an ML approach to model travel mode choice [1, 6–11]. These authors have generally reported improvements in the prediction performance of ML approaches compared to MNL-based models. Recently, Wang et al. established an empirical benchmark by using 86 ML models to predict travel mode choice based on a 2017 U.S. national household travel survey dataset [12]. The authors found that ensemble models such as boosting, bagging, and random forest models exhibit performances superior to those of all other ML methods, including deep neural networks. However, due to the black-box nature of ML models, the authors could not explain the prediction results, making it difficult to find a suitable explanation for the relationship between the input variables and travel mode choices.

Several studies have performed additional analyses of the prediction results to complement the evaluation of performance. Wang and Ross proposed an extreme gradient boosting (XGB) model for predicting travel mode choice [1]. Using a relatively comprehensive dataset, the authors measured the relative importance of variables in the training process of the XGB and estimated the importance of correlated variables that cannot be explained using the MNL model. Hagenauer and Helbich measured the permutation-based importance of variables in predicting the choice of each travel mode, and their result showed that the critical variables varied with the predicted travel modes [7]. Lee et al. developed a choice model for alternatives related to autonomous vehicles using a gradient boosting machine (GBM) [10]. They measured the partial dependence (PD), which captures the marginal effects of attributes representing the relationship between the input variables and predicted output. Although the above researchers who conducted these three studies tried to explain the prediction results of their ML models with several meaningful interpretations, there is room for improvement by the application of various interpretation methods to reveal details of the characteristics of travel behavior.

In this study, model-agnostic interpretation methods were applied to explain the prediction results of ML models concerning mode choice behavior. XGB, random forest (RF), and artificial neural network (ANN) models were employed to predict travel mode choices from national household travel survey (NHTS) data in Seoul. Trip- and tour-related attributes were extracted from the NHTS data to construct the variable set. The tour refers to interconnected trips (i.e., trip chain) during a day. This dataset is enriched with traffic analysis zone (TAZ)-level spatial information. The performance of the models was evaluated regarding their prediction of each travel mode. Then, the best-performed XGB prediction results were analyzed to reveal choice behavior for urban travel modes. In doing so, two crucial issues were addressed, which are difficult to investigate using a conventional MNL model, i.e., (i) how each variable interacted with other variables and (ii) how the variable related to the probability of travel mode choice.

The remainder of this paper is organized as follows. In Section 2, the dataset and data-processing procedure applied in this study are described. Then, the ML models and model-agnostic interpretation methods are discussed in detail. In Section 3, performance evaluation of the ML models and interpretation of the XGB prediction results are presented. Finally, concluding remarks and future research directions are presented in Section 4.

2. Materials and Methods

2.1. Data Descriptions. The primary source of data for this study was a 2016 NHTS dataset in the Seoul, Korea [13]. These data included individual travel diaries that recorded every daily trip taken, with multiple trips on a given day expressed as a trip chain. The chained trips were divided by their trip purpose and established the major travel modes of the trip's purpose. For example, a person who uses the subway to go to work must first access the subway station on foot and then use the subway. In this case, the two chained trips, walking and subway, are combined into one subway trip as the primary travel mode. Walking is considered a primary travel mode only if it is used as the sole travel mode, but not as a means to access another travel mode. Seoul operates a public transit unified fare system for buses and subways, whereby charges are levied as if the person is using a single travel mode when transferring between these two forms of public transit. Therefore, this study makes no distinction between a bus and a subway, whereby the chained trips of a bus and subway with a transfer are considered to be one trip by public transit.

Table 1 describes the variables included in the travel mode choice model. Four categories of variables are used to train and test the mode choice model. Trip-related, tour-related, and individual attributes are extracted from the NHTS data, and built environment attributes are obtained from national spatial data [14] and population census [15] in Korea. The departure and arrival locations of NHTS data are recorded in the TAZ unit, which is within a radius of about 1 km; thereby, the NHTS data are merged with built environment attributes according to TAZ. The dependent variable is for primary travel modes: car, bike, transit, and walking. A single mode, which is assumed, is used for an entire tour because 89.9% of the respondents in the NHTS data used the one primary travel mode rather than a combination of modes. Trip-related attributes are extracted from single or sequential individual trips. The duration of an activity is calculated by the difference between the arrival time on the previous trip and the departure time on the next trip. The duration of activity on the last trip (i.e., the return trip home) is calculated by the difference between the arrival time of the last trip and the departure time of the first trip. Travel time includes in-vehicle and out-of-vehicle time, such

TABLE 1: Description of the independent and dependent variables.

Variable name	Explanation	Data types
Travel mode	Chosen travel mode for the trip (dependent variable): 1 = car, 2 = bike, 3 = transit, and 4 = walking	Categorical
Trip-related attributes		
Activity duration	Duration of the activity	Numeric
Travel time	Travel time of the trip	Numeric
Departure time	1 = the trip occurs in the morning or evening peak hours (8 A.M.–10 A.M. or 5 P.M.–7 P.M.); 0 = otherwise	Dummy
Trip type	Context of the trip: 1 = home-based work (HBW); 2 = home-based others (HBO); 3 = non-home-based others (NHBO); 4 = return home (RH)	Categorical
Tour-related attributes		
Sum of activity duration	Sum of activity duration during a day excluding the last trip	Numeric
Sum of travel time	Sum of travel time during a day	Numeric
Number of trips	Number of trips that occurred during a day	Categorical
Tour type	Context of the tour: 1 = home-work-home (HWH); 2 = home-other-home (HOH); 3 = home-work-other-home (HOWH)	Categorical
Individual attributes		
Age	Age of the traveller in years	Numeric
Gender	1 = the traveller is male; 2 = the traveller is female	Dummy
Car owner	1 = the household of traveller owns a car; 0 = otherwise	Dummy
Driver's license	1 = the traveller has a driver's license; 0 = otherwise	Dummy
Income	Monthly household income of the traveller (million KRW): low = income < 5; high = income ≥ 5	Dummy
Built environment attributes		
Land use in D: residential	The ratio of residential area to the total area at D in TAZ unit	Numeric
Land use in D: commercial	The ratio of commercial area to the total area at D in the TAZ unit.	Numeric
Population density at D	Density of the population (people/km^2) at the destination in the TAZ unit	Numeric
Number of workers at D	Number of workers at the destination in the TAZ unit	Numeric
Number of bus stops at D	Number of bus stops at the destination in the TAZ unit	Numeric
Number of subway stops at D	Number of subway stops at the destination in the TAZ unit	Numeric

Note. D = destination of a trip; 1,000 KRW = 0.84 USD.

as waiting time and access time. Departure time is divided into peak and nonpeak categories. Trip type is defined by the characteristics of the origin and destination, such as home, work, or other places. Tour-related attributes are extracted from all the trips of individuals during a single day. The sum of activity durations of trips is calculated, excluding the last trip, and the sum of travel time and the number of trips includes all the trips made during a day. Tour types are defined by the combination of trip types included in a tour. The Home-Other-Home (HOH) type includes the tour with more than three trips (e.g., H-O-O-H). Individual attributes include age, gender, car owner, driver's license, and income, and all of those attributes are directly collected in the NHTS data. Built environment attributes describe the spatial characteristics of a trip's destination (D). The variables for land use are defined as the ratio of a residential or commercial area to the total area. Population density, number of workers, number of bus stops, and number of subway stops are also used to characterize the destination in the TAZ unit. Although travel cost is an important variable in the travel mode choice, the NHTS data used in this study did not include the respondents' travel cost such as fuel cost, parking cost, and transit fares. Therefore, the effect of travel cost does not consider in the analysis like other studies using the

NHTS data [1, 7, 8]. After a data-cleaning process, in which the trips were removed with very long activity duration and travel time, a total of 172,889 trips taken by 76,190 individuals were used. 75% of the NHTS data was used for training and 25% of those data for the test.

Table 2 shows the descriptive statistics of the variables. The distribution of the travel mode is imbalanced in that trips by walking, transit, car, and bike are 43.7%, 35.3%, 18.5%, and 2.5%, respectively. The mean activity duration is 490.2 minutes, which is slightly longer than the standard working time of eight hours, and the mean travel time of each trip is 21.7 minutes. The number of trips during a peak time is comparable to the number of trips at a nonpeak time. In terms of trip type, the percentage of HBW, HBO, NHBO, and RH are 31.8%, 16.7%, 4.8%, and 46.7%, respectively, indicating that more than 20% of noncommuting trips are included in the data. The sum of activity duration and the sum of travel time have a mean value of 509.2 minutes and 51.6 minutes, respectively. While 70.9% of travelers make two trips during a day, 29.1% make more than three trips. The people who made more than three trips may have tour types of HOH or HOWH, which are 27.0% and 21.4% of total tours, respectively. The percentages of females, car owners, driver's licenses, and those with a high income are

TABLE 2: Descriptive statistics of the variables.

Variable name	Category	%	Mean	Standard deviation
Travel mode	Car	18.5		
	Bike	2.5		
	Transit	35.3		
	Walking	43.7		
Trip-related attributes				
Activity duration (min)			490.2	251.1
Travel time (min)			21.7	15.9
Departure time	Peak	50.6		
	Nonpeak	49.4		
Trip type	HBW	31.8		
	HBO	16.7		
	NHBO	4.8		
	RH	46.7		
Tour-related attributes				
Sum of activity duration (min)			509.2	235.6
Sum of travel time (min)			51.6	33.4
Number of trips	2	70.9		
	3	10.5		
	4	16.4		
	5	1.4		
	6	0.8		
Tour type	HWH	51.6		
	HOH	27.0		
	HOWH	21.4		
Individual attributes				
Age			44.6	20.0
Gender	Female	51.7		
	Male	48.3		
Car owner	Yes	72.0		
	No	28.0		
Driver's license	Yes	54.7		
	No	45.3		
Income	High	33.0		
	Low	67.0		
Built environment attribute				
Land use in D: residential			0.49	0.20
Land use in D: commercial			0.29	0.20
Population density at D			42,862	11,771
Number of workers at D			32,787	75,271
Number of bus stops at D			125.2	85.5
Number of subway stops at D			1.0	1.2

Note. D = destination of a trip.

51.7%, 72.0%, 54.7%, and 33.0%, respectively. While the car owner indicates whether the household owns a private car, the driver's license indicates whether the individual owns a driver's license. The descriptive statistics of built environment attributes are also presented in Table 2.

2.2. Machine Learning Model for Predicting Travel Mode Choice.
Three ML models, XGB, RF, and ANN, were applied to predict travel mode choices. Given a set of values of the input variable, the model predicts the probability that a specific travel mode will be chosen. To account for class imbalance, weight to the data instance is applied in inverse proportion to the frequency distribution of each class, and those class-specific weights are commonly used to train ML models. A hyperparameter is a parameter that controls the training process of the ML model. Since the hyperparameter affects the speed and quality of the training process, hyperparameter tuning is an essential task for evaluating an ML model's performance. The major hyperparameters of

each ML model were tuned using a grid search technique based on 4-fold cross-validation. A comparable degree of a set of hyperparameter combinations is considered for each ML model.

2.2.1. Random Forest.

The decision tree is a popular ML model due to its ability to capture complex structures in the data, although it suffers from an overfitting problem. To address this issue, ensemble models have been proposed. The RF [16] is a tree-based ensemble method related to the bagging approach, which averages noisy but approximately unbiased models to reduce the variance. An ensemble of independent trees on a random subset of a training dataset with randomly selected variables can achieve better generalized performance [9, 17]. The RF has also shown promising performance for predicting travel mode choice in previous studies [7, 8]. There are four significant hyperparameters used to tune the learning process of an RF model: the number of trees, the number of variables to split in each node, the maximum depth of each tree, which determines the model complexity of each tree, and the data-sampling rate used for training each tree. The RF model is implemented using the "ranger" package in R [18].

2.2.2. Extreme Gradient Boosting Model.

The GBM is another tree-based ensemble method that has been successfully used to predict travel mode choice [1, 10]. Unlike the RF, the GBM builds a sequence of the low-depth decision tree, where each tree is trained to put more weight on the incorrect prediction of the previous trees [19]. The results of all the estimated trees collectively determine the result of the ensemble model. To implement GBM, an eXtreme Gradient Boost (XGB) proposed by Chen et al. [20] is employed. XGB is an efficient algorithm for constructing boosted trees using regularization terms and parallel processing. The five major hyper parameters of XGB are tuned, including the learning rate, maximum depth of each tree, number of variables considered in each tree, number of samples considered in each tree, and minimum value of the sum of instance weight of a node. The XGB model is implemented using the "xgboost" package in R [20].

2.2.3. Artificial Neural Network.

The ANN is a widely used ML model for the training classification model. The promising performance of ANN rather than MNL for modeling travel mode choice has been reported in previous studies [6, 7]. A multilayer perceptron (MLP) is a conventional neural network including an input layer, one or more hidden layers, and an output layer. Nonlinear relationships in the data can be naturally captured by the MLP since it iteratively adjusts the weights and biases between neurons' interactions in multiple layers [21]. This study adopts an MLP with a single hidden layer, and a standard backpropagation algorithm with a decay term was used to train the MLP. The number of neurons in the hidden layer and a decay term are tuned. The ANN model is implemented using the "nnet" package in R [22].

2.3. Model-Agnostic Interpretation Methods.

Interpretability is defined as the degree of understanding the cause of prediction [23]. Traditional interpretable models, such as logistic regression and decision tree, sacrifice prediction performance due to a simple model structure that improves interpretability. Recently, model-agnostic interpretation methods have been applied to make machine learning interpretable. Those interpretation methods commonly measure changes in prediction performance according to changes in the value of input variables. By doing so, the marginal effect of the variables is estimated to deduce the importance and interaction of variables. Also, the complex relationship between the input and outcome can be estimated. The target of the interpretation methods is divided into two perspectives: the entire model behavior (i.e., global interpretability) and a single prediction (i.e., local interpretability) [24]. This study focuses on the former by applying three model-agnostic interpretation methods.

2.3.1. Permutation-Based Variable Importance.

When values of a variable are permutated so that their relationship with the predicted outcome is broken, the prediction error will increase. By calculating the increases in the model's prediction error, the importance of the variable is obtained. This study measures the importance based on the algorithm proposed by Fisher et al. [25]. The permutation-based variable importance can naturally consider all interactions with other variables (i.e., the sum of main and cross effects) by permutation. Therefore, highly correlated variables also can be directly interpreted. For the input variable matrix $\mathbf{X}$, the original error (e^{orig}) of the ML model ($\widehat{f}$) is estimated by the defined loss function (L) between the predicted value ($\widehat{f}(\mathbf{X})$) and the true value (y), as in equation (1). Then, the input matrix, including the permutated variable j ($\mathbf{X}^{\mathrm{perm}_j}$) is used to compute the permutated error (e^{perm_j}), and the importance of variable j (VIMP_j) is calculated by ($e^{\mathrm{perm}_j}/e^{\mathrm{orig}}$), as shown in equation (2):

$$e^{\mathrm{orig}} = L(y, \widehat{f}(\mathbf{X})), \tag{1}$$

$$\mathrm{VIMP}_j = \frac{L\left(y, \widehat{f}\left(\mathbf{X}^{\mathrm{perm}_j}\right)\right)}{e^{\mathrm{orig}}} = \frac{e^{\mathrm{perm}_j}}{e^{\mathrm{orig}}}. \tag{2}$$

To measure the importance of the multiclass classification, the balanced accuracy of each travel mode (see equation (3)) is used as a L between the predicted value and the true value:

$$\mathrm{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}},$$

$$\mathrm{Sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{3}$$

$$\mathrm{Balanced\ accuracy} = \frac{\mathrm{Specificity} + \mathrm{Sensitivity}}{2},$$

where TN, FN, TP, and FP are the true negative, false negative, true positive, and false positive, respectively. Compared with the accuracy, the balanced accuracy can serve as a better judge of performance for the imbalanced

classification problem where the difference in the number of negative and positive samples for each class is large [26]. The balanced accuracy in this study also measures the prediction performance of the ML model.

2.3.2. Variable Interaction.

When variables are correlated, the effect of one variable depends on the value of other variables. The change in the prediction error also can be used to measure those correlations (i.e., variable interaction). Friedman's H-statistic is used to estimate the strength of the variable interaction quantitatively. This measurement indicates how much the variation in the prediction depends on the interaction of the variables [27]. The marginal effect of a variable on the model's prediction is represented by the partial dependence (PD) function, as in

$$\mathrm{PD}_j(x_j) = \frac{1}{n} \sum_{i=1}^{n} \widehat{f}(x_j, x_{-j}^{(i)}),$$

$$\mathrm{PD}_{jk}(x_j, x_k) = \frac{1}{n} \sum_{i=1}^{n} \widehat{f}(x_j, x_k, x_{-j-k}^{(i)}), \tag{4}$$

where $\mathrm{PD}_j(x_j)$ is the PD function of a single variable j, $\mathrm{PD}_{jk}(x_j, x_k)$ is the 2-way PD function of two variables j and k, n is the total number of data points, i is a certain data point used to estimate the marginal effect, x_j and x_k are the variables used to calculate the marginal effects, and x_{-j} and x_{-j-k} are the other variables used in the ML model ($\widehat{f}$). Mathematically, the interaction between variables j and k (i.e., two-way interaction) is estimated as in equation (5), and the interaction between variable j and any other variables (i.e., total interaction) is estimated as in equation (6) [28]:

$$H_{jk}^2 = \frac{\sum_{i=1}^{n} \left[\mathrm{PD}_{jk}\left(x_j^{(i)}, x_k^{(i)}\right) - \mathrm{PD}_j\left(x_j^{(i)}\right) - \mathrm{PD}_k\left(x_k^{(i)}\right) \right]^2}{\sum_{i=1}^{n} \mathrm{PD}_{jk}^2\left(x_j^{(i)}, x_k^{(i)}\right)}, \tag{5}$$

$$H_j^2 = \frac{\sum_{i=1}^{n} \left[\widehat{f}\left(x^{(i)}\right) - \mathrm{PD}_j\left(x_j^{(i)}\right) - \mathrm{PD}_{-j}\left(x_{-j}^{(i)}\right) \right]^2}{\sum_{i=1}^{n} \widehat{f}^2\left(x^{(i)}\right)}, \tag{6}$$

where $\mathrm{PD}_{-j}(x_{-j}^{(i)})$ is the PD function that depends on all variables except the jth variable. While the two-way interaction in equation (5) indicates the amount of the variance explained by the interaction between the two variables x_j and x_k among the variance of the output of the PD, the total interaction in equation (6) indicates the amount of the variance explained by the interaction between variables x_j and any other variable x_{-j} among the variance of the output of the entire function [28]. Therefore, if the H-statistic is zero, there is no interaction at all, and if all the effect of variables is applied as an interaction, the statistic would be one. When the H-statistic is larger than one, the interpretation would be difficult. In the case of two-way interaction, this can happen when the variance of two-way interaction is larger than the variance of the two-dimensional PD In the

case of total interaction, this can happen when the variance of interaction between one variable and other variables is larger than the variance of the ML model.

2.3.3. Accumulated Local Effect.

The promising performance of the ML model suggests that complex relationships exist between the input variables and predicted outcome in the real data, which may be nonlinear or polynomial. To represent these relationships, the ALE value was used, which shows the changes in the probability of a travel mode choice by the specific value (or category) of a variable. Generally, the marginal effect of the variables can be obtained using the PD function [10, 17]. However, the PD function assumes that the variables are not correlated with each other, which is unrealistic in real data. When the variables are highly correlated, the PD function includes unrealistic data when averaging the prediction results, which can substantially bias the estimated effect of the variable [28]. To address this issue, the accumulated local effect (ALE) is used, which is the unbiased alternative to PD [29]. The value of ALE can be interpreted as the main effect of the variable at a specific value compared to the average prediction value of the data. The ALE plots can depict any relationship, whether linear, monotonic, or more complex, between a variable and the predicted outcome. The ALE calculates the change in prediction results by replacing the target variable with grid values z. The average change in prediction is the effect for a specific interval, and its effect accumulates across all intervals as [29]

$$\widehat{f}_{j,\mathrm{ALE}}(x_j) = \sum_{k=1}^{k_j^K(x)} \frac{1}{n_j(k)} \sum_{i:\, x_j^{(i)} \in N_j(k)} \left[f\left(z_{k,j}^K, x_{-j}^{(i)}\right) - f\left(z_{k-1,j}^K, x_{-j}^{(i)}\right) \right], \tag{7}$$

where $z_{k,j}^K$ is the partition of the minimum and maximum of x_j into K interval and $k_j^K(x) = k$ if $x \in (z_{k-1}^K, z_{k,j}^K]$, the average effects of all instances within an interval ($N_j(k)$) are calculated by dividing the sum of the difference of the prediction, i.e., $\sum_{i:\, x_j^{(i)} \in N_j(k)} [f(z_{k,j}^K, x_{-j}^{(i)}) - f(z_{k-1,j}^K, x_{-j}^{(i)})]$, by the number of instances in this interval ($n_j(k)$). The ALE is centered on having a zero mean, as shown in

$$\widehat{f}_{j,\mathrm{ALE,cent}}(x_j) = \widehat{f}_{j,\mathrm{ALE}}(x_j) - \frac{1}{n} \sum_{i=1}^{n} \widehat{f}_{j,\mathrm{ALE}}(x_j^{(i)}). \tag{8}$$

While the intervals can be defined by the distribution of the numeric variables, the intervals for the categorical variables are determined by the similarity of categories since the categorical variables do not have a natural order. The similarity of the two categories is calculated by the sum of distances over the other variables. While the distance between the target category and other numeric variables is calculated by Kolmogorov–Smirnov distance, the distance between target category and other categorical variables is calculated by the relative frequency tables. More details are described in [28].

3. Results and Discussion

3.1. Prediction Performance. Since the travel modes are imbalanced, the prediction performance of the RF, XGB, and ANN models are evaluated using three metrics: specificity, sensitivity, and balanced accuracy, as shown in equation (3). Table 3 compares the prediction performances of the three models. Overall, the RF and XGB models exhibit better performance than the ANN model. Although class-specific weight was applied for training the ML models, all models show poor performance for the prediction of bike choice that is minority class (i.e., 2.5% of total). The performance of the XGB is comparable to that of RF and exhibited better performance for some travel modes and metrics. Compared with the RF, XGB shows slightly lower performance for predicting the choices of car and bike but shows better performance for predicting the choice of transit and walking. For all travel modes, the XGB shows the best performance for all metrics.

The number of FN explains the low sensitivity of the XGB for minor classes (i.e., car and bike). For example, in the case of car, the number of FN is 2,635, including 1,489 transit, followed by 1,111 walking and 35 bike. This result indicates that consideration of trip- and tour-related attributes cannot successfully identify the choice of car and public transit. This may be because the competitiveness of public transit (i.e., relative travel time for given OD) in Seoul is as high as that of cars [30]. The FN caused by walking indicates that car and walking share some travel characteristics. This result can be explained by travel patterns in Seoul where short-distance driving (i.e., trips of 5 km or less) represent 44% of all car driving [31]. The short-distance driving can indicate similar travel time to walking trip. In the case of bike, the number of FN is 754, including 401 walking, 219 transit, and 134 car. It also indicates that the travel characteristics of walking are similar to those of bike, such as travel time and trip type. To develop an understanding of mode choice behavior, the prediction results of the best-performing XGB model were analyzed using three model-agnostic interpretation methods in the following section.

3.2. Variable Importance. The permutation-based variable importance was measured based on the XGB model. Since decision makers have different objectives and application plans for each travel mode, the importance was measured for each travel mode. Figure 1 shows box plots of the importance of the top ten variables for each travel mode, which was calculated from 50 simulations to consider the randomness introduced by the permutation. Since this importance considers both the main and cross effects of a variable, it cannot be interpreted as the main effect of variables like the coefficient of MNL.

Although some variables are commonly important in predicting all mode choice, the ranking of other variables is somewhat different. Travel time and activity duration are important for all travel modes, and their influence is more significant on a tour level than on a trip level. The result can explain the recent success of the tour-based model in travel demand forecasting, compared with the trip-based model [32, 33]. While age, travel time, and activity duration commonly rank highly in importance among all travel modes, car owner, land use, and number of trips only influence a specific travel mode. This implies that policy-making needs to be carried out by focusing on different factors for each travel mode, based on the mode-specific analysis.

Regarding car, age is the most important variable in determining choice, which may indicate the varying preference for comfort and value of time by age [34]. Car ownership, of course, is the second important variable for the choice of a car. Two tour-related attributes, the sum of travel time and the sum of activity duration, are more critical than two corresponding trip-related attributes, travel time and activity duration.

Regarding bike, the small number of positive samples of bike results in a higher variance of importance than other travel modes. Low performance of the XGB may cause those variances, and the proposed box plot is useful in the case of those high variances. Similar to car, the age, sum of travel time, and sum of activity duration rank highly in terms of importance for bike, followed by gender. Unlike other travel modes, two land-use variables show considerable importance, indicating that land use affecting accessibility and mobility would influence the use of bikes [35].

Transit and walking present similar patterns of importance ranking. Both travel time for a trip and tour are important variables for the choice of transit and walking, followed by age and activity duration. As for walking, travel time is a dominant factor since only a short distance can be travelled, and, as for transit, travel time is a critical criterion for determining competitiveness over car and bike [36]. Both travel modes are significantly affected by the number of trips on tour and how the number of trips affects the choice of transit and walking is discussed in a later section using ALE.

3.3. Variable Interaction. Variable interaction was measured for each travel mode using the H-statistic. As shown in equations (5) and (6), the variable interaction can be divided into two cases, i.e., total interaction and two-way interaction. The left side of Figure 2 shows the total interaction of the top ten variables for the choice of each travel mode. Further investigation of total interaction is conducted by two-way interaction, as shown in the right side of Figure 2.

Regarding car, age, sum of activity duration, activity duration, sum of travel time, and travel time are found to have high interaction with other variables. The two-way interactions also indicate that their high interactions are caused mainly within them. This result reveals that their effects on prediction consist of main and significant cross effects, which cause the high variable importance of those variables (see Figure 1). For example, interaction strength between the sum of travel time and travel time is 0.37, which means 37% of the effect of those two variables on the prediction comes through the interaction. On the contrary, the car owner has a low interaction but high importance, indicating that the effect of the car owner appears mainly as

TABLE 3: Comparison of prediction performances of the ML models.

Travel modes	Specificity			Sensitivity			Balanced accuracy		
	ANN	RF	XGB	ANN	RF	XGB	ANN	RF	XGB
Car	0.806	0.881	**0.920**	0.752	**0.713**	0.670	0.779	**0.797**	0.795
Bike	0.985	0.990	**0.993**	0.116	**0.338**	0.291	0.550	**0.664**	0.642
Transit	0.879	**0.887**	0.883	0.515	0.639	**0.744**	0.697	0.763	**0.813**
Walking	0.819	0.834	**0.850**	0.739	0.826	**0.856**	0.779	0.830	**0.853**
All	0.882	0.909	**0.923**	0.647	0.727	**0.768**	0.764	0.818	**0.845**

Note. ANN = artificial neural network; RF = random forest; XGB = extreme gradient boosting.
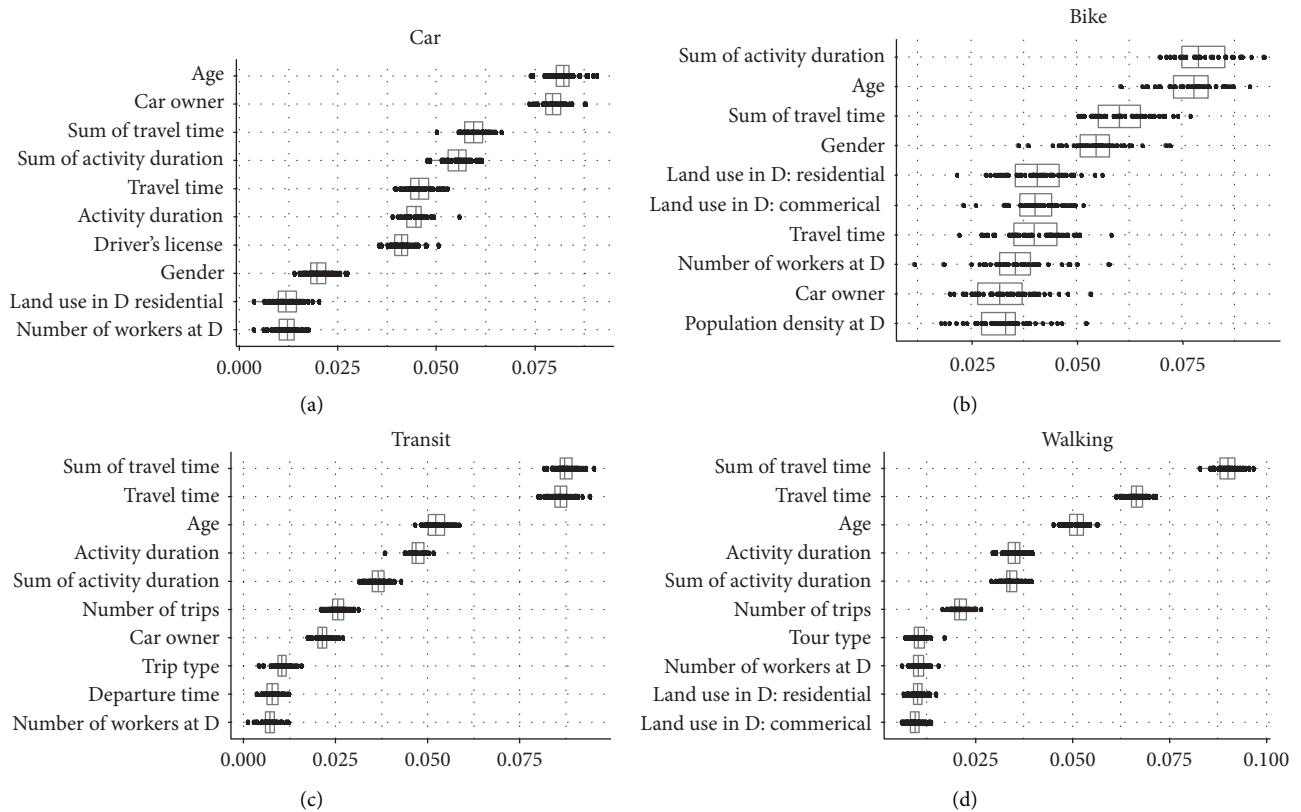
FIGURE 1: Permutation-based variable importance for each travel mode choice: (a) car, (b) bike, (c) transit, and (d) walking.
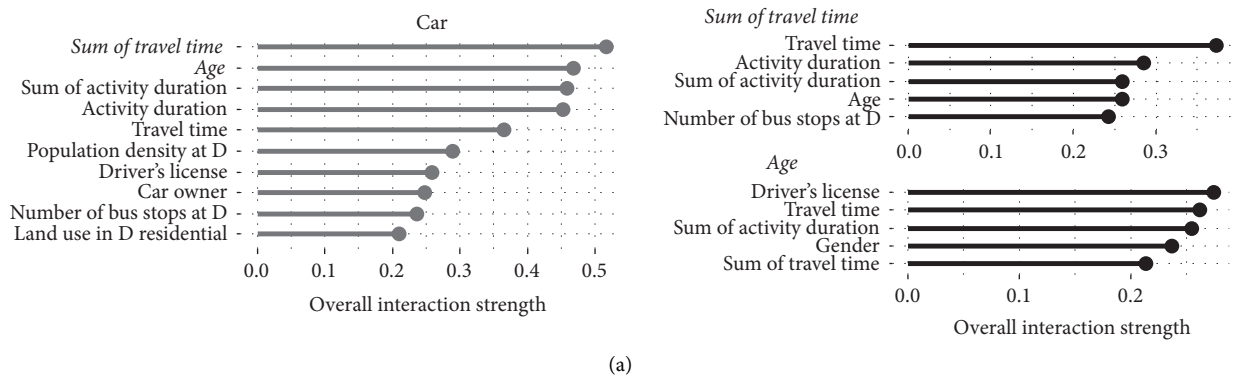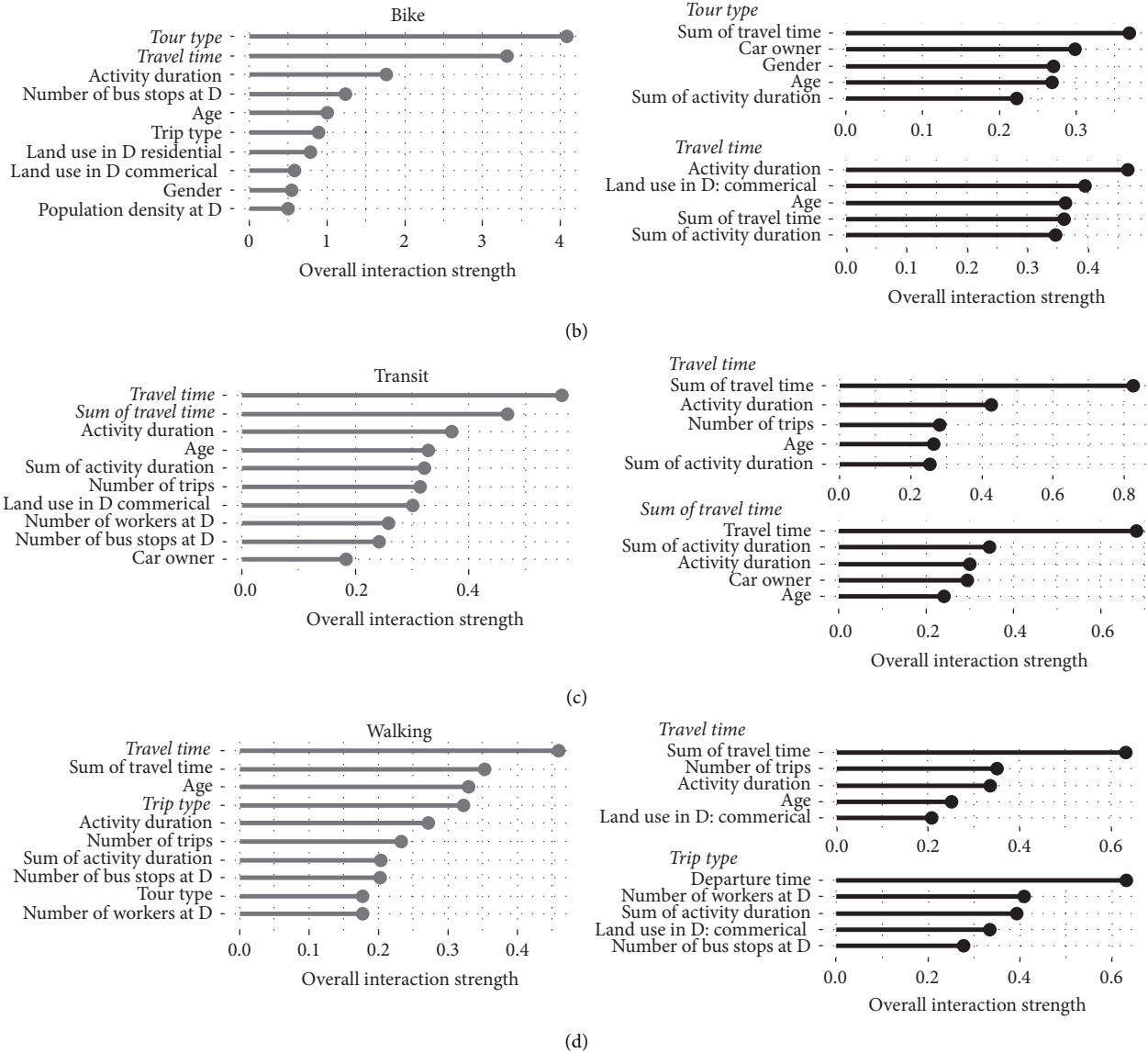

(a)

FIGURE 2: Continued.

FIGURE 2: Variable interactions between one and other variables and between two variables, in predicting travel mode choice of (a) car, (b) bike, (c) transit, and (d) walking.

the main effect. Since the car owner indicates whether the household owns a private car, it can have low interaction with other individual attributes. Age shows the highest interaction with driver's license due to age restrictions on driver's license, although the car owner is more important for the choice of car than the driver's license. High interaction between age and gender indicates that gender, which is not top ten important variables, affects prediction mainly through interaction (i.e., cross effect).

Regarding bike, total interactions are higher than one, indicating that the variance of total interaction is larger than the variance of the ML model. This result can be caused by the low specificity (0.291) of the XGB model to bike choice, of which the changes in the value of a variable cannot thoroughly explain the changes in the class probability of bike. Therefore, it is difficult to extract significant meaning to

the interpretation of the total interaction of the bike. Although the two-way interactions for bike have interaction strength smaller than one, significant interpretation is still challenging due to the result of total interaction.

Transit and walking show similar patterns of variable interaction, just like variable importance. Travel time, activity duration, and age have high total interactions for both travel modes. The two-way interaction of travel time for transit and walking choice indicate that, like car choice, the effects of travel time that are of high importance are derived from the significant cross effects among travel time, activity duration, and age. The number of trips is found to have high interaction with travel time for both transit and walking choice. This reveals that the use of transit and walking can be determined by a combination of travel time and the number of trips on tour, which measure the travel fatigue. Unlike
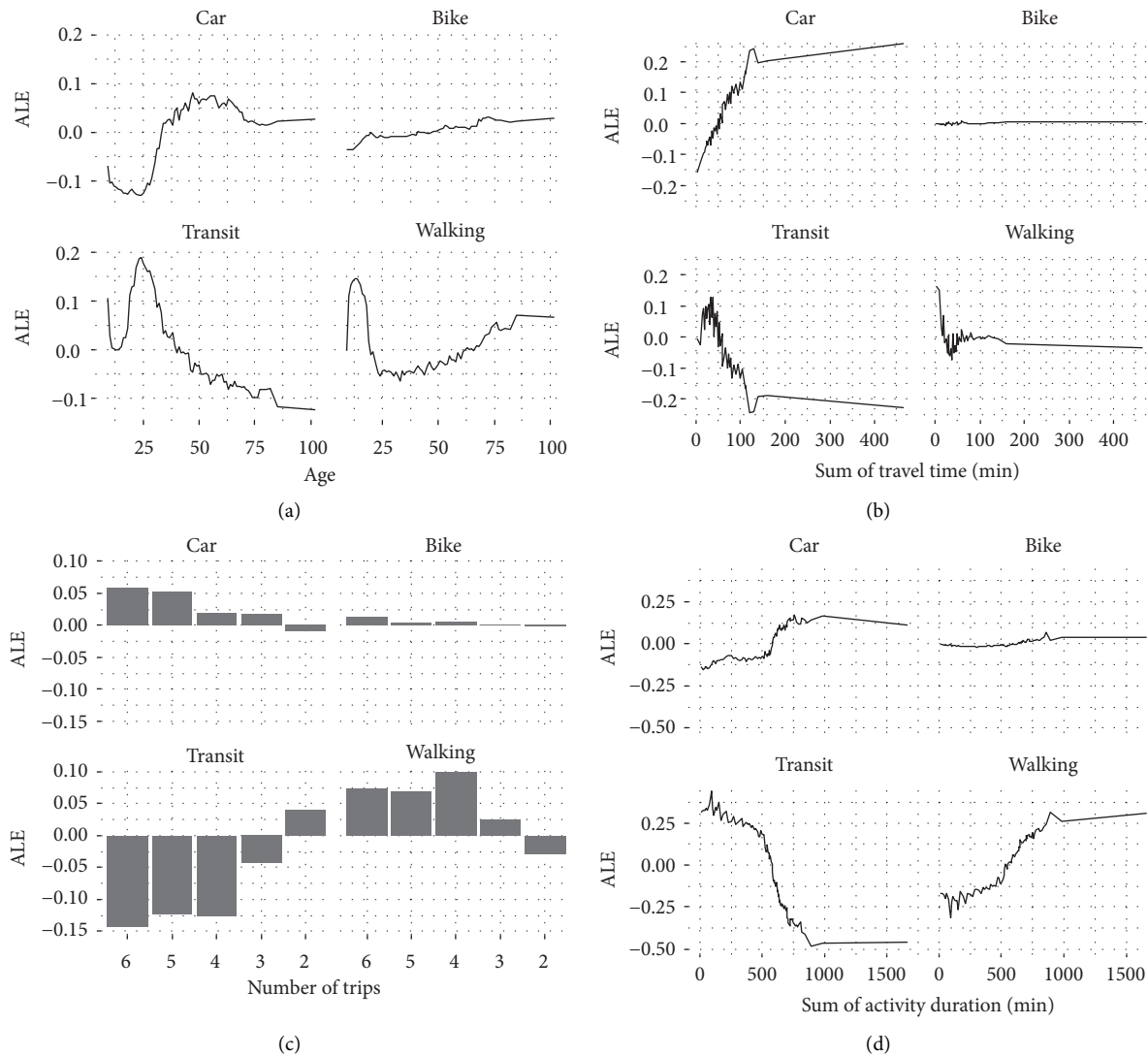
FIGURE 3: The ALE values of variables for predicting each travel mode: (a) car, (b) bike, (c) transit, and (d) walking.

other variables presenting a similar pattern for transit and walking, trip type has high total interaction for walking, while low total interaction for transit. Further investigation by two-way interactions shows that trip type is highly correlated with departure time, number of workers at D, activity duration, and land use, which are closely related to trip purpose [37]. The fact that walking includes both trip type and tour type in ten important variables also supports this result. This may be because the choice of walking is significantly linked to eating out and social/recreational trips or going school trip of the student [38].

3.4. Relationship between Variable and Travel Mode Choice.
Although the variable importance and interaction tell us the magnitude of the importance and interactions, they do not

present how they work. Based on variable importance and interaction, the significant variables are selected for further investigation by the ALE plots, as in Figure 3. While variable importance measures the total effect, including the cross and main effect, the value of ALE measures the main effect of a variable at a specific value (or specific category) on the prediction. Therefore, as shown in Figure 3, age that has a relatively high interaction and importance, and the number of trips that have relatively low interaction and importance can have a similar magnitude of ALE.

Age represents notable patterns of ALE for each travel mode. The choice probability of car gradually increases as age increases from the 20s to 60s, and decreases after the mid-60s, which may suggest a relationship between physical ability or social status and choice of car [38, 39]. The choice of bike gradually increases as age increases, but the

difference is tiny. This result is caused by the lack of explanatory power of the XGB model in predicting bike choice. The choice of transit rises steadily until the mid-20s when people graduate from university and then decreases. Teenagers and older people in the study prefer walking as a travel mode more than those of other ages. The choice of walking, after reaching a high in the teenage years, declines toward the 30s and subsequently increases gradually. The peak ALE value of 0.15 among 14 year olds means the probability of walking being chosen is 15% higher for people who are 14 years old than the average age. The above nonlinear relationship between age and travel mode choice is valuable information that cannot be observed from conventional MNL assuming a linear relationship.

The ALE of the categorical variable is also calculated. As the number of trips increases, the choice probability of car and walking increases while the choice probability of transit decreases. This indicates that the number of trips would be a barrier to transit use as it is generally more burdensome to undertake multistop tours [40]. Meanwhile, a large number of trips would include trips of a relatively short distance, such as leisure and shipping trips, so the choice probability of walking would have increased. For bike, near-zero ALE appears, similar to age.

As the sum of travel time and the sum of activity duration increase, the tendency to choose car increases, while the tendency to choose transit decreases. Specifically, when the sum of travel time is more than 50 minutes, the choice probability of car and transit is symmetrical, and this pattern is also observed in the ALE of the sum of activity duration. This result intuitively indicates that car and transit are alternative to each other, depending on travel time and activity duration. When the sum of travel time and activity duration increases, the choice probability of car increases while those of transit decreases. The tendency to use walking as a travel mode decreases as the sum of travel time increases and is maintained after a slight rebound. This rebound may be related to the interaction between the number of trips and the travel time since a large number of trips would include more short-distance trips. People who perform activities for more than 500 minutes a day tend to use a car and walk more than transit. Considering that eight hours are regarded as the average number of working hours, the sum of activity duration is also an indicator for an additional trip activity after/before work, which would be short-distance trip. Therefore, the choice probability of walking continues to increase as the sum of activity duration increases.

4. Conclusions

This paper proposed interpretable ML approaches to predicting and analyzing travel mode choice. The XGB model performed best in the prediction of travel mode choice relative to the RF and ANN models. Understanding the

decisions made by the XGB model is valuable both for improving prediction performance and providing insight to the practitioner. The three model-agnostic interpretation methods, i.e., permutation-based variable importance, H-statistic-based variable interaction, and ALE, were applied to investigate the influence of variables in predicting travel mode choices. These methods uncovered the correlated and nonlinear relationships between the behavioral attributes and travel mode choice.

Some interesting findings were highlighted by the results of three interpretation methods. The results of variable importance revealed that age, travel time, and activity duration have high importance for all travel modes. The interactions of those variables explained that such high importance is caused by large cross effects among those variables. These interrelated aspects of the significant variables revealed why the ML model considering the complex relationship of variables outperforms the traditional statistical models in predicting travel mode choice, as reported in the previous studies [1, 6–8]. Also, the tour-related attributes showed high interaction and importance for the choice of all travel modes, indicating that the tour-based analysis is necessary for mode choice, as reported in a modern travel demand forecasting model [41]. These findings regarding the complexity of mode choice emphasized the need to shift from the existing MNL model to a flexible ML model. The varying importance of some variables such as the car owner, tour type, land use, and number of trips according to travel mode indicated that mode-specific analysis should be conducted for targeting each travel mode. For example, to accurately predict the walking trips in the location, trip purpose-related attributes such as land use and activity duration should be collected. The ALE successfully represented the nonlinear relationship between the variables and the change in the choice probability of each travel mode, which is difficult to derive from a conventional MNL. The ALE intuitively showed the alternative patterns of travel mode through the symmetric patterns between travel modes. These results revealed the detailed modal shift patterns according to the behavior attributes such as age and the sum of travel time, which could be used to guide how to divide people into subgroups for predicting travel demand of each mode.

In future research, a proposed interpretation method is needed to extend a more in-depth and broader understanding of travel behavior. Bivariate ALE can be applied to represent the cross effect between variables that separated from the main effect, and it can enrich the explanation of variable interaction. Comparing the interpretation results of ML models with an advanced parametric model, such as a mixed logit model, would also be valuable to validate the model further. Deep learning models [11, 42] are reasonable alternatives for the XGB and RF and the proposed model-agnostic interpretation methods can still available for those models. Local interpretation methods such as local interpretable model-agnostic explanations (LIME) and Shapley

additive explanations (SHAP) can contribute to better representation of the heterogeneity of individuals and groups [43, 44], which has also been a critical subject of behavior analysis. Although this study only considers a single primary mode due to the regional travel pattern, a tour-based mode choice model considering the exact combination of modes has been recently proposed to consider the dynamics among trips within the tour [45, 46]. Applying the proposed ML and interpretation methods to those complex modeling tasks would be meaningful future research in the regions with a high rate of multimodal trips.

Data Availability

All data used in this study are available from the author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] F. Wang and C. L. Ross, "Machine learning travel mode choices: comparing the performance of an extreme gradient boosting model with a multinomial logit model," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 47, pp. 35–45, 2018.

[2] E. J. Miller, "Modeling the demand for new transportation services and technologies," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2658, no. 1, pp. 1–7, 2017.

[3] D. A. Hensher, W. H. Greene, and Z. Li, "Embedding risk attitude and decision weights in non-linear logit to accommodate time variability in the value of expected travel time savings," *Transportation Research Part B: Methodological*, vol. 45, no. 7, pp. 954–972, 2011.

[4] C. R. Bhat, "Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling," *Transportation Research Part A: Policy and Practice*, vol. 32, no. 7, pp. 495–507, 1998.

[5] K. E. Train, *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, UK, 2009.

[6] D. Lee, S. Derrible, and F. C. Pereira, "Comparison of four types of artificial neural network and a multinomial logit model for travel mode choice modeling," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 49, pp. 101–112, 2018.

[7] J. Hagenauer and M. Helbich, "A comparative study of machine learning classifiers for modeling travel mode choice," *Expert Systems with Applications*, vol. 78, pp. 273–282, 2017.

[8] S. Rasouli and H. J. P. Timmermans, "Using ensembles of decision trees to predict transport mode choice decisions: effects on predictive success and uncertainty estimates," in *Proceedings of the 17th International Conference of Hong Kong Society for Transportation Studies*, Hong Kong, China, 2012.

[9] A. Lhéritier, M. Bocamazo, T. Delahaye, and R. Acuna-Agost, "Airline itinerary choice modeling using machine learning," *Journal of Choice Modelling*, vol. 31, pp. 198–209, 2019.

[10] D. Lee, J. Mulrow, C. J. Haboucha, S. Derrible, and Y. Shiftan, "Attitudes on autonomous vehicle adoption using interpretable gradient boosting machine," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, no. 11, pp. 865–878, 2019.

[11] Z. Zhang, C. Ji, Y. Wang, and Y. Yang, "A customized deep neural network approach to investigate travel mode choice with interpretable utility information," *Journal of Advanced Transportation*, vol. 2020, Article ID 5364252, 11 pages, 2020.

[12] S. Wang, B. Mo, and J. Zhao, "Predicting travel mode choice with 86 machine learning classifiers: an empirical benchmark study," in *Proceedings of the 99th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, 2020.

[13] KTDB (Korea Transport Database), *Household Travel Survey Data*, KTDB, Republic of Korea, 2016, https://www.ktdb.go.kr/eng/contents.do?key=263.

[14] MOLIT (2016), http://www.nsdi.go.kr/lxportal/?menuno=3085.

[15] KOSTAT (2016), http://kostat.go.kr/portal/eng/index.action.

[16] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[17] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Science & Business Media, Berlin, Germany, 2009.

[18] M. N. Wright and A. Ziegler, "Ranger: a fast implementation of random forests for high dimensional data in C++ and R," *Journal of Statistical Software*, vol. 77, pp. 1–17, 2017.

[19] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[20] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, and H. Cho, "Xgboost: extreme gradient boosting," *R Package Version 0.4-2*, pp. 1–4, 2015.

[21] E.-J. Kim, H.-C. Park, S.-Y. Kho, and D.-K. Kim, "A hybrid approach based on variational mode decomposition for analyzing and predicting urban travel speed," *Journal of Advanced Transportation*, vol. 2019, Article ID 3958127, 12 pages, 2019.

[22] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, Springer, New York, NY, USA, 4th edition, 2002.

[23] T. Miller, "Explanation in artificial intelligence: insights from the social sciences," *Artificial Intelligence*, vol. 267, pp. 1–38, 2019.

[24] A. Adadi and M. Berrada, "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[25] A. Fisher, C. Rudin, and F. Dominici, "All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously," *Journal of Machine Learning Research*, vol. 20, pp. 1–81, 2019.

[26] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," in *Proceedings of the 2010 20th International Conference on Pattern Recognition*, Istanbul, Turkey, 2010.

[27] J. H. Friedman and B. E. Popescu, "Predictive learning via rule ensembles," *The Annals of Applied Statistics*, vol. 2, no. 3, pp. 916–954, 2008.

[28] C. Molnar, "Interpretable machine learning: a guide for making black box models explainable," 2019, https://christophm.github.io/interpretable-ml-book/.

[29] D. W. Apley and J. Zhu, "Visualizing the effects of predictor variables in black box supervised learning models," 2016, https://arxiv.org/abs/1612.08468.

[30] H. Lee, H.-C. Park, S.-Y. Kho, and D.-K. Kim, "Assessing transit competitiveness in Seoul considering actual transit travel times based on smart card data," *Journal of Transport Geography*, vol. 80, Article ID 102546, 2019.

[31] S. Lee, "A study on decreasing methods of short-distance auto trips in Seoul," Report: 2004-R-08, Seoul Development Institute, Seoul, Republic of Korea, 2004.

[32] J. Freedman, J. Castiglione, and B. Charlton, "Analysis of new starts project by using tour-based model of San Francisco, California," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1981, no. 1, pp. 24–33, 2006.

[33] M. Bradley, J. L. Bowman, and B. Griesenbeck, "SACSIM: an applied activity-based model system with fine-level spatial and temporal resolution," *Journal of Choice Modelling*, vol. 3, no. 1, pp. 5–31, 2010.

[34] J. Ha, S. Lee, and J. Ko, "Unraveling the impact of travel time, cost, and transit burdens on commute mode choice for different income and age groups," *Transportation Research Part A: Policy and Practice*, vol. 141, pp. 147–166, 2020.

[35] A. Faghih-Imani, N. Eluru, A. M. El-Geneidy, M. Rabbat, and U. Haq, "How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal," *Journal of Transport Geography*, vol. 41, pp. 306–314, 2014.

[36] T. A. Arentze and E. J. E. Molin, "Travelers' preferences in multimodal networks: design and results of a comprehensive series of choice experiments," *Transportation Research Part A: Policy and Practice*, vol. 58, pp. 15–28, 2013.

[37] E.-J. Kim, Y. Kim, and D.-K. Kim, "Interpretable machine learning models for estimating trip purpose in smart card data," *Municipal Engineer*, Institution of Civil Engineers, London, UK, 2020.

[38] S. Kim and G. F. Ulfarsson, "Curbing automobile use for sustainable transportation: analysis of mode choice on short home-based trips," *Transportation*, vol. 35, no. 6, pp. 723–737, 2008.

[39] J. Scheiner, "Social inequalities in travel behaviour: trip distances in the context of residential self-selection and lifestyles," *Journal of Transport Geography*, vol. 18, no. 6, pp. 679–690, 2010.

[40] L. Yang, Q. Shen, and Z. Li, "Comparing travel mode and trip chain choices between holidays and weekdays," *Transportation Research Part A: Policy and Practice*, vol. 91, pp. 273–285, 2016.

[41] J. Castiglione, M. Bradley, and J. Gliebe, *Activity-based Travel Demand Models: A Primer*, Transportation Research Board, Washington, DC, USA, 2015.

[42] E.-J. Kim, H.-C. Park, S.-W. Ham, S.-Y. Kho, and D.-K. Kim, "Extracting vehicle trajectories using unmanned aerial vehicles in congested traffic conditions," *Journal of Advanced Transportation*, vol. 2019, Article ID 9060797, 16 pages, 2019.

[43] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems*, pp. 4765–4774, MIT Press, Cambridge, MA, USA, 2017.

[44] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *Proceedings of the 22nd ACMSIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, San Francisco, CA, USA, 2016.

[45] M. S. Hasnine and K. Nurul Habib, "Tour-based mode choice modelling as the core of an activity-based travel demand modelling framework: a review of state-of-the-art," *Transport Reviews*, vol. 41, no. 1, pp. 5–26, 2021.

[46] P. Vovsha, J. E. Hicks, G. Vyas et al., "Combinatorial tour mode choice," in *Proceedings of the 96th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, 2017.