Hindawi
Journal of Advanced Transportation
Volume 2023, Article ID 2206625, 23 pages
https://doi.org/10.1155/2023/2206625

Research Article

Introducing Autonomous Shuttle Services Based on Travel Patterns for the Elderly

Eunbi Kang, Sunmin Park, Younghoon Seo, and Hyungjoo Kim

Laboratory of Intelligent Transportation System, Advanced Institute of Convergence Technology, Suwon 16229, Republic of Korea

Correspondence should be addressed to Hyungjoo Kim; hyungjoo@snu.ac.kr

Received 26 November 2022; Revised 27 December 2022; Accepted 25 March 2023; Published 24 April 2023

Academic Editor: Ziyuan Pu

The transportation-disadvantaged population is rapidly increasing through aging; moreover, the mobility of the elderly has increased due to life expectancy improvements and lifestyle changes. To respond to these changes, a customized mobility service specifically tailored to the travel patterns of the elderly was developed in this study. In particular, an autonomous shuttle service that can improve the operational efficiency and the accessibility of existing transportation methods was proposed by considering the travel characteristics of elderly people, who mainly travel short distances. To this end, a study was conducted in Seongnam City, Republic of Korea, where mobility support services for the elderly are insufficient. Using smart card data, the elderly were classified according to their travel patterns, with autonomous shuttle routes suggested for each travel purpose. We derived four clusters via the Gaussian mixture model clustering, with travel purposes classified according to the spatiotemporal travel patterns. Finally, the major routes for each travel purpose were selected, and feasible road paths for autonomous shuttle operation were suggested using the concept of ODD (operational design domain). This holistic methodology is expected to contribute to the development of autonomous mobility services for the elderly and for the establishment of welfare policies based on the smart card data.

1. Introduction

According to the Act on Promotion of the Transportation Convenience of Mobility Disadvantaged Persons, a transportation disadvantaged is a person who has difficulty in moving in daily life. Such individuals include the elderly, disabled, pregnant women, people with infants, and children. In 2020, the number of people with transportation difficulties in Korea was 15.4 million, corresponding to 29.7% of the total population and an increase of ~182,000 compared to the 2019 statistics [1]. Approximately, 8.5 million people (or 55.2%) of those with transportation disabilities are elderly, and their population is steadily increasing every year. According to this trend, by 2025, the elderly population in South Korea will account for 20.3% of the total population, and Korean society is expected to become "super-aged." In addition, the mobility of the elderly is gradually being improved by changes in life expectancy and lifestyle [2]. The number of elderly drivers' license holders has increased by ~700,000, with an average of 11.2% per year [3]. To ensure the safety of pedestrians and vehicles, local governments are encouraging elderly drivers to return their licenses voluntarily. Accordingly, various measures to promote the use of public transportation among the elderly are being implemented, but these measures are not sustainable. Currently, mobility services for the elderly are classified into methods that utilize existing public transportation modes and special transportation services. In the case of existing public transportation, low-floor buses and subway free rides are being adopted to support the elderly. However, low-floor buses are difficult to supply owing to replacement costs, and free rides are a major factor in the deficit of subway systems [4]. Moreover, as special transportation services require qualifications such as disability level, it is difficult to use it easily. Therefore, additional measures are required to support the mobility and to provide long-term stable services for the elderly.

Such mobility support services must focus on improving the accessibility of transportation. Accessibility is primarily lacking in the first mile (i.e., the access trip from origin to public transit service) and the last mile (i.e., the egress trip to the destination after using public transportation), thus the first and the last miles are major causes of difficulty when elderly use public transportation. Recently, autonomous mobility services which could serve as feeders are being proposed to solve such accessibility problems [5, 6]. In particular, when providing mobility services using autonomous driving, the labor costs can be significantly reduced, thereby reducing operating losses and the burden on public finances. As driver costs account for more than 60% of the total public transportation operating costs in Gyeonggi-do, this aspect should be considered a top priority when providing mobility services [7]. Thus, autonomous driving can offer higher-frequency and more accessible services, which is expected to improve the convenience of the elderly.

In this paper, we introduce a data-driven approach to plan an autonomous customized shuttle for the elderly to sustainably enhance their mobility. Customized mobility services aim to provide efficient mobility support to groups of passengers with similar travel patterns [8, 9]. Since it is essential to identify groups with similar travel patterns and demands for a sustainable customized service operation, we applied a clustering methodology that is widely used for classifying travel patterns. In particular, considering the travel characteristics of the elderly, who mainly travel short distances, we planned a local circulation shuttle service, and classified the elderly according to their boarding and alighting patterns. As a result, by designing an autonomous shuttle that reflects the travel patterns of the elderly who get on and off in similar areas, it is expected that the elderly will be able to move efficiently and conveniently by using the shuttle within their main activity space.

The remainder of this paper is organized as follows: Section 2 describes the previous research studies on travel pattern analysis and autonomous mobility services. Section 3 discusses a methodology of clustering travel patterns and planning routes for each travel type. Section 4 presents the case study of Seongnam and analysis results derived under the research methodology. Finally, Section 5 presents the conclusions and limitations of this study, as well as topics for future research studies.

2. Literature Review

2.1. Travel Pattern Analysis Using Smart Card Data. Smart card data include information, such as card ID, boarding time, boarding point, and user type, thus allowing spatiotemporal travel histories of passengers to be extracted [10, 11]. As the development of public transportation systems facilitates large-scale smart card data collection and loading, various studies using these data are being conducted. In particular, the smart card makes continuous tracking possible even at the level of individual users, and studies to evaluate the adequacy of public transportation services [12, 13] and to analyze and estimate travel patterns [14, 15] are underway.

El Mahrsi et al. [16] used the unigram model to classify users according to boarding times. To this end, a dataset was constructed by counting the number of trips by day and time of the week for each individual user, with stops clustered according to the travel time using a Poisson mixed model. Devillaine et al. [17] used smart card data from Santiago and Gatineau to classify travel purposes according to the travel date/time and boarding/alighting locations and compared the analysis results for each city. In their study, the purpose of travel was classified under "work," "school," "home," and "other," and the standards for defining the purpose of each travel were presented (e.g., for "school": bus card type is "student" or "minor," and "bus stop is in an education area"). Ma et al. [18] analyzed the travel patterns and the regularity of transit riders. The DBSCAN algorithm was employed to integrate multiple travel routes for individual users into one common route, and the k-means++ algorithm was employed to cluster the users according to the number of travel days, ride times, routes, and stops. The users were classified into five types according to the degree of regularity; and it was confirmed that the higher the regularity, the greater the number of trips and travel time. Lee and Lee [19] proposed a method of detecting travel patterns that were recognized as different journeys as a single flow travel pattern, using smart card data. After measuring the proximity between the boarding and alighting points of each travel using the agglomerative flow clustering algorithm, the single flow cluster grouping process was repeated to categorize similar travel patterns. Existing smart card-based travel pattern analysis studies have primarily been conducted for general users, but with the aging trend, research studies are also underway to identify the travel patterns targeting the elderly. Liu et al. [20] classified elderly people with similar travel patterns according to the seasonal variability. The dataset included the ratio of the number of trips per month and the average of the number of trips for elderly people who frequently use public transportation. The k-means++ algorithm was used for classification, and the variability of types with similar monthly travel patterns was analyzed. Shao et al. [21] compared the weekday and weekend spatiotemporal travel patterns of the elderly and general users by utilizing the travel distance, frequency, and ride time. Clustering was performed for spatially adjacent stops, and the travel network was analyzed using the number of travels between the boarding and alighting clusters.

Until now, studies regarding the analysis and clustering of travel patterns using smart card data have been continually conducted for general users and the elderly. However, previous studies have been limited to travel pattern analysis, and detailed studies implementing the analysis results have been lacking. Hence, this study analyzes the travel patterns of the elderly and presents a corresponding methodology for implementing new mobility services.

2.2. Implementation of Autonomous Mobility Service. Alongside research studies into autonomous driving technology, research studies on autonomous mobility combining public and shared transportation is also being conducted in various ways. Yoon et al. [22] defined several autonomous public transport service types that could be introduced in Korea depending on the vehicle specifications and operating conditions such as autonomous driving shuttles, trunk lines/bus rapid transit, demand responsive transit (DRT), and branch lines. In this study, when introducing autonomous public transportation, improved traffic safety, urban space efficiency, labor cost, and service quality were presented as expected effects. Lim et al. [23] analyzed the impacts of introducing autonomous shared transportation services into the market according to the city's modal share in Korea through simulation. When autonomous shared transportation services are provided, the share of cars will be reduced and the effect of reducing traffic congestion will be large, which means that a shared autonomous driving system can positively impact the city traffic operation.

Recently, research studies on the design and demonstration of autonomous driving public transport and shared transport services has also been conducted. Gurumurthy and Kockelman [24] designed an autonomous DRT shuttle in Orlando using cell phone data. Matching between individual trips was performed using the origin, destination, and the departure time of the mobile phone-based travel data. In travel matching, the service was planned by adopting the approach that overlapping routes exist between the boarding and alighting points, even if the departure and arrival locations differ. Dandl et al. [25] proposed an autonomous driving shuttle for work and identified areas not well connected to the public transportation to replace private commutes. Thus, the operation area was defined as an area in which the travel time of public transportation exceeded the driving time of a passenger car at peak times, and the usage rate of public transportation was below 30%. Shuttle stops were set around residential areas or nearby arterial roads, and the shuttle was planned to assign all requested stops to the quickest route using the Dijkstra algorithm. Rehrl and Zankl [26] introduced and demonstrated the first- and last-mile autonomous driving shuttles in suburban areas of Koppl. The test route was designed as a round-trip route of approximately 2.8 km, including four bus stops. Lin et al. [27] also conducted a demonstration of an autonomous shuttle service within a campus. To implement the shuttle service, the demonstration area was selected based on the student travel and proximity to major facilities; and candidate routes were proposed by considering the route length, V2I communication availability, and speed limit. The final route was determined as a route offering high safety and frequent contact with passengers and pedestrians, making it easy to evaluate the performance of the autonomous shuttle.

Various studies have been conducted regarding the design and demonstration of autonomous mobility services; however, studies that systematically reflect the characteristics of autonomous driving in the design stage are insufficient. Therefore, it is necessary to deduce factors to be considered when utilizing autonomous driving as well as to perform research studies that fully reflect these factors.

2.3. Implications. In smart card data-based studies, the analysis and classification of travel patterns have been performed for general users and the elderly, and the k-means clustering algorithm has been primarily applied to classify according to the temporal travel patterns by counting the number of rides per unit of time. Previous studies have focused on the analysis stages, and insufficient attention has been paid to improving existing public transportation services or designing new service plans. By reviewing the literature regarding the introduction of autonomous mobility services, it was confirmed that the studies focusing on the elderly are insufficient, and it is necessary to consider the users' travel patterns and autonomous driving characteristics in an integrated manner. Moreover, several studies related to the travel behaviors and demand of the elderly have suggested that they should be divided into heterogeneous subgroups and should be analyzed for improving the public transportation systems and for implementing customized mobility services [28–30].

Thus, this study aims to propose a comprehensive framework for introducing a customized autonomous shuttle for the elderly. To this end, we classify elderly people with similar travel patterns and designed possible routes based on the spatiotemporal travel patterns of each group. Afterwards, by analyzing the feasibility of autonomous driving according to the operational design domain (ODD) factors, major routes that autonomous shuttles can operate were presented. The contribution of this study is summarized as follows.

(1) To provide customized shuttles for the elderly with similar travel patterns, we constructed a 20-day trip-chain data of individual elderly and proposed a methodology by applying GMM to cluster the passengers according to their boarding and alighting patterns.

(2) In addition, to explore the specific travel networks for each type, origin-destination pairs and boarding time patterns were analyzed. Unlike the frequency analysis for each origin and destination point, this approach analyzes the pair in which the origin and destination are linked and makes us plan routes that reflect the actual travel flow of the elderly.

(3) By conducting a case study of Seongnam and providing the real road paths for autonomous driving, a series of steps for introducing a customized autonomous shuttle from travel pattern clustering-route planning-ODD analysis were proposed. This framework can contribute to laying the foundation for the government and policymakers to better understand and introduce the customized autonomous shuttle services.

3. Methodology

3.1. Data Preprocessing and Dataset Construction. The smart card data contained 44 columns, of which the 16 columns (including anonymized card ID, user type code, boarding area code, and boarding time) required for the analysis were extracted. A card ID is a unique ID possessed by an individual smart card user, allowing users to be identified; and the user type code can identify the specific user types (e.g., disabled and student). Descriptions of the other items used in this study can be found in Table 1. The user type and the mode of transportation codes were used to extract the elderly people's bus travels. In addition, to extract only normal travel records, abnormal travel records related to boarding and alighting were examined. Abnormal travel records can be largely divided into two types [31]. (1) The first is a case in which columns to be used for analysis contain missing values, and (2) the second is a case in which when two or more columns are logically inconsistent. Since this study uses both boarding and alighting records, the first case where the time and area code of the alighting are missing was considered an abnormal travel record. The second case corresponds to the case where the alighting time is earlier than the boarding time and where the boarding and alighting stops are the same. Since the final destination cannot be confirmed, it cannot be used for travel pattern analysis when the boarding and alighting stops are the same. Therefore, all rows with missing values related to alighting were dropped, those in which the alighting time was earlier than the boarding time, and those where the boarding and alighting stops were the same were dropped sequentially.

A customized mobility service aims to provide mobility support to a group of passengers with similar travel patterns. Thus, we developed a methodology for identifying the similar spatial travel pattern of the elderly by counting the number of boardings and alightings for each dong (neighborhood unit). To estimate the original travel patterns and purposes, we only retained the boarding area of the first trip and the alighting area of the last trip among the single-purpose trips linked by the transfers. According to the integrated transfer fare system for the metropolitan area, transfer travel was applicable only to those who boarded within 1 h in the period 21 : 00–07 : 00 and within 30 min otherwise.

The dataset was constructed with high dimensionality, owing to the number of dongs. Dimensionality reduction was applied to solve these problems as high-dimensional datasets can reduce the computational efficiency and performance of the classification models.

The representative dimensionality reduction algorithms include principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE). PCA performs dimensionality reduction via linear transformations; and in the case of nonlinear datasets, this can reduce the model performance [32]. In contrast, t-SNE can be applied to both linear and nonlinear data, and it differs from existing dimension-reduction techniques by solving the crowding problem (i.e., points far enough away in a high dimension cannot be implemented in a low dimension owing to the limited space). Previous studies that compared the clustering results after applying several dimensionality reduction algorithms have confirmed that t-SNE yields superior clustering results compared to others [32, 33]. Therefore, in this study, t-SNE was used for dimensionality reduction before clustering.

t-SNE compares the similarity between high- and low-dimensional data and calculates this similarity using the Euclidean distance. The similarity p_{ij} of the high-dimensional data x_i, x_j is calculated as follows:

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n} \text{ where } p_{j|i} = \frac{\exp\left(-\left\|x_i - x_j\right\|^2/2\sigma_i^2\right)}{\sum_{k\neq i}\exp\left(-\left\|x_i - x_k\right\|^2/2\sigma_i^2\right)}, \quad (1)$$

where $p_{j|i}$ is the conditional probability that x_i will select x_j as a neighbor and σ_i denotes the variance of the normal distribution in which x_i is centered. Moreover, in t-SNE, $p_{i|i} = 0$ and $p_{ij} = p_{ji}$ are defined.

The similarity q_{ij} of the low-dimensional data y_i, y_j for comparison is expressed as follows:

$$q_{ij} = \frac{\left(1 + \left\|y_i - y_j\right\|^2\right)^{-1}}{\sum_{i\neq k}\left(1 + \left\|y_i - y_k\right\|^2\right)^{-1}} \text{ where } q_{i|i} \triangleq 0, \quad (2)$$

t-SNE can significantly classify q_{ij} with respect to the distance between data by adopting a t-distribution with one degree of freedom instead of a normal distribution.

We quantified the mapping to a low-dimensional data via the similarity of P and Q calculated from high- and low-dimensional joint probability distributions and utilized the Kullback–Leibler divergence. The corresponding cost function is as follows:

$$C = \mathrm{KL}(\mathrm{P}\|\mathrm{Q}) = \sum_i \sum_j p_{\mathrm{ij}} \log \frac{p_{\mathrm{ij}}}{q_{\mathrm{ij}}}. \quad (3)$$

When C is minimized, it means that the two distributions are similar.

To increase the similarity between the low- and high-dimensional data, learning was performed using a gradient descent. When $Y^{(t)}$ is the t-th low-dimensional dataset, η is the learning rate, and $\alpha(t)$ is the t-th momentum, then the process of updating the low-dimensional data using the gradient descent is as follows:

$$Y^{(t)} = Y^{(t-1)} + \eta \frac{\delta \mathrm{C}}{\delta Y} + \alpha(t)\left(Y^{(t-1)} - Y^{(t-2)}\right). \quad (4)$$

3.2. Travel Pattern Clustering. The clustering algorithms are the unsupervised learning methods that define the similarity between the data and accordingly cluster them into the most similar groups. These algorithms can be largely divided into hierarchical, density-based, and partitioning methods [34, 35]. Hierarchical clustering is a method of grouping the most similar clusters according to the distance between all clusters. As the number of data increases, the computational burden rapidly increases, so large datasets cannot be effectively processed in this manner [36]. Density-based clustering methods such as DBSCAN perform clustering on densely distributed data. Therefore, DBSCAN has a fast

Table 1: Summary of selected columns in smart card data.

Column	Description
City, county, and district code	City, county, and district codes governing the transportation
Transportation classification code	Codes assigned to bus types and subways
Card ID	Anonymized card ID
Boarding time	User's ride time
Departure time	User's drop-off time
Last departure time	User's last drop-off time
Difference time	Time taken between previous and current transactions
Number of transfers	Number of rides within the transfer time after alighting
Transaction ID	Travel record ID
User type code	User type code for children and general public.
Distance	Distance traveled by public transport
Travel time	Time taken in transit
Boarding stop ID	Stop ID where user boards
Alighting stop ID	Stop ID where user alights
Boarding area code	Area code where user boards
Alighting area code	Area code where user alights

processing speed but is unsuitable for datasets with large density variations [37]. Partitioning clustering improves the partition by splitting data according to a predefined number of clusters (k) and repeatedly relocating data points. As the partitioning clustering algorithm has a linear time complexity with respect to the data size, it can efficiently process large datasets [38]. Consequently, it has been applied to various research fields and is widely used in studies involving large smart-card datasets [39, 40]. The most typical partitioning clustering algorithm, the k-means algorithm, rapidly determines cluster membership by minimizing the distance between the central point and the other data points in the cluster. However, the k-means algorithm is sensitive to outliers, and it struggles to identify clusters that are nonspherical or that exhibit different sizes and densities [41, 42]. In contrast, the Gaussian mixture model (GMM) can identify clusters with different sizes or densities and more general shapes; thus, it is more flexible than that of the k-means clustering [43, 44]. In many cases, the GMM algorithm shows a higher classification accuracy than that of the k-means clustering [42, 45] and model-based clustering methods are suitable for modeling similar travel patterns [36, 40].

We explored the characteristics of each algorithm and compared them to the dataset used in this study. First of all, the hierarchical method is inefficient for large datasets in this study because of its high computational complexity. As shown in Figure 1, the density of the data is not uniform, and the data are distributed in clusters of various shapes. In the case of the density-based DBSCAN, relatively low-density parts can be treated as noise. In addition, the k-means algorithm tends to create spherical clusters of equal volume; thus, it may be too restrictive for pattern analysis and for performing clustering flexibly [46]. On the contrary, GMM is suitable for large sizes of data because of its low computational complexity and enables the formation of clusters of various sizes and shapes, making it an algorithm suitable for the dataset in this study. Furthermore, it is a method that is often used in recent travel pattern clustering studies [47–49].

GMM clusters data assuming that the data are combined with several Gaussian distributions. The parameters constituting the GMM include the number of normal distributions K, the mean μ, and the covariance Σ for each normal distribution, and the mixing coefficient w, which indicates the probability that the distribution is selected. GMM estimates parameters through the expectation-maximization (EM) algorithm. The expectation step calculates the probability that the data were generated from a random distribution to derive the probability γ and that the data belonged to the cluster. The maximization step improves the distribution by updating parameters μ, Σ, and w using γ. GMM repeats the expectation and maximization steps a given number of times or until the log-likelihood (used for parameter evaluation) is maximized, and this process can be confirmed in Algorithm 1. After learning the parameters of each normal distribution, the k^{th} normal distribution with the highest $\gamma(z_{\text{nk}})$ is defined as the probability distribution generating x_n, and in terms of clustering, it is classified as the k^{th} cluster.

The Akaike information criterion (AIC) and Bayesian information criterion (BIC) are the most popular criteria for evaluating GMM [50, 51]. Both the AIC and BIC evaluate the criteria that can explain the distribution of data well, and they calculate the score by imposing a penalty on the complexity of the model (i.e., the number of normal distributions). In particular, the BIC imposes a higher penalty on the model complexity than on AIC and is widely adopted because it can select the most effective model at a low computational cost [52]. Therefore, in this study, the number of normal distributions in GMM, that is, the optimal number of clusters was derived by using BIC.

$$\text{BIC} = -2\log L + K\log n. \tag{5}$$

Where

L maximized value of the likelihood function

K: number of parameters to be estimated

n: number of observations

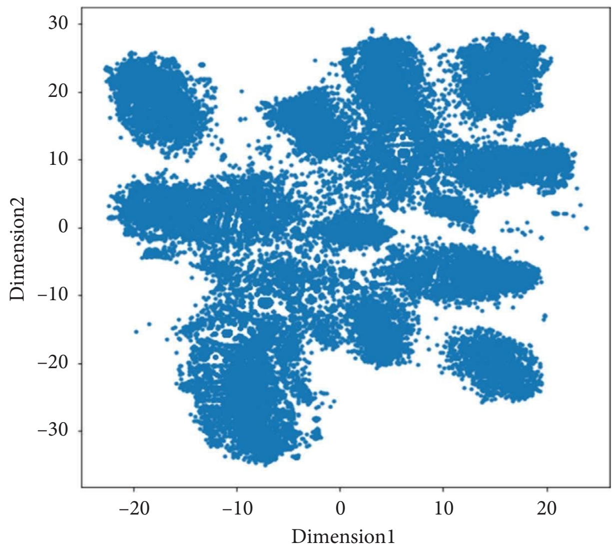


Figure 1: Two-dimensional representation of data using t-SNE.

3.3. Travel Pattern Analysis. After applying GMM, we analyzed the spatiotemporal travel characteristics for each cluster and defined the trip purpose accordingly. The spatial distribution of travel was analyzed by using the combination of boarding and alighting stops (i.e., the O-D pair) to reflect the real travel flow of the elderly. By deriving the top ten O-D pairs in terms of the travel volume, each O-D pair was visualized in the target area, with the travels assigned to a 500 m grid to identify the influence zone of the boarding and alighting stops [53].

Then, to verify the temporal distribution of O-D pairs for each cluster, the number of travels in each time interval for each O-D pair was aggregated from 05 : 00 to 24 : 00 (the operating times of public transportation services). In many studies, the temporal dimension is used to estimate the travel purpose of public transportation users, and it was found that the boarding time reflects the travel purpose well [54, 55]. According to a household travel survey in the metropolitan area, shopping and leisure trips for the elderly primarily occur between 09 : 00 and 12 : 00, and most return trips between 12 : 00 and 18 : 00, suggesting that the boarding time varies depending on the purpose of the trip. Therefore, in this study, the travel purpose of the cluster was estimated using the hourly travel volume of the O-D pair, as calculated using the boarding time, and the general travel characteristics of the elderly population. In addition, we utilized the distribution of the major facilities that the elderly make trips to within a 1 km radius of the top ten O-D pairs [2].

3.4. Autonomous Shuttle Service Routes Based on Trip Purpose. The autonomous shuttle service supports the first- and the last-mile connections to improve the accessibility of the existing public transportation using a low-speed autonomous vehicle [22]. As discussed earlier, autonomous shuttles are attracting attention as an option for the first- and last-mile services for the elderly [56]. In particular, because personal mobility such as electric kickboards or bicycles, currently popular as the last-mile transportation, are not appropriate for users with limited physical abilities, thus, autonomous shuttles are more suitable modes for the elderly. In addition, autonomous shuttles offer excellent operational efficiency and economic feasibility through the frequent operation of short-distance routes [57, 58]. Therefore, it is expected that the last mile problems can be effectively solved if autonomous shuttle services are operated on round trips from the public transportation nodes to the destinations frequently used by the elderly.

In this study, to consider the travel characteristics of the elderly who mainly board near the subway stations and to increase the connectivity with existing the public transportation services, the bus stops near the stations were selected as the starting points of the shuttle route. Moreover, to provide the first- and last-mile services, a circular route was constructed using the destinations of the top ten O-D pairs as stops and endpoints.

Routes were planned based on actual roads, and we used components of the autonomous driving ODD to verify whether autonomous shuttles could be operated. ODD provides information for defining autonomous driving system functions; this includes road types, geographic conditions, speed ranges, and other domain constraints [59, 60]. We used conditions based on the ODD components defined in the previous studies, but certain components were revised to incorporate the travel characteristics of the elderly and the collected data. The ODD components are largely divided into geometric, operational, and environmental factors and we collected the corresponding data for each major route. The detailed criteria for determining autonomous driving is shown in Table 2.

4. Applications and Result Analysis

4.1. Study Site. As shown in Figure 2Seongnam-si is one of the 31 local governments in Gyeonggi-do, South Korea, which has an area of 141.7 km^2, and contains three districts

Input: $X = \{x_1, \ldots, x_N\}$ where $x_i \in \mathbb{R}^D$
Output: $w = \{w_1, \ldots, w_K\}$ where $w_i \in \mathbb{R}$
$\mu = \{\mu_1, \ldots, \mu_K\}$ where $\mu_i \in \mathbb{R}^D$
$\Sigma = \{\Sigma_1, \ldots, \Sigma_K\}$ where $\Sigma_i \in \mathbb{R}^{K\times K}$
Randomly initialize μ, Σ, w
for $t = 1: T$
  for $n = 1: N$ (E-Step)
    for $k = 1: K$
      $\gamma(z_{\mathrm{nk}}) = (w_j N(x_n|\mu_j, \Sigma_j)/\sum_{j=1}^{K} w_j N(x_n|\mu_j, \Sigma_j))$
    end
  end
  for $k = 1: K$ (M-Step)
    $\mu_k = (\sum_{i=1}^{K}\gamma(z_{\mathrm{nk}})x_n/\sum_{i=1}^{K}\gamma(z_{nk}))$
    $\Sigma_k = (\sum_{i=1}^{K}\gamma(z_{\mathrm{nk}})(x_n - \mu_k)(x_n - \mu_k)^T/\sum_{i=1}^{K}\gamma(z_{nk}))$
    $\pi_k = 1/N\sum_{i=1}^{K}\gamma(z_{\mathrm{nk}})$
  end
end
X: dataset; x_i: i^{th} data point; N: size of dataset
w_i: Probability that the i^{th} distribution will be selected
μ_i: Mean of the i^{th} distribution
Σ_i: Covariance matrix of the i^{th} distribution
$\gamma(z_{\mathrm{nk}})$: Probability that a data point x_n belongs to the k^{th} distribution

Algorithm 1:Expectation-maximization (EM) algorithm for GMM.

(gu) and 44 legal-status neighborhoods (dong). Its population is approximately 920,000 as of 2022, with the elderly population accounting for 15.3% of the total population. In addition, Jungwon-gu and Sujeong-gu in Seongnam rank within the top five major travel districts for the elderly in Gyeonggi-do (Korea Transportation Institute View-T 3.0). However, Seongnam currently lacks mobility support services for the elderly, with no plans in place to introduce such services other than special transportation by 2024 [61]. In addition, Pangyo Techno Valley, Seongnam, has been selected as an autonomous driving testbed and is currently piloting an autonomous shuttle service. Therefore, we selected Seongnam as the mobility services for the elderly are insufficient and an autonomous shuttle can be easily introduced.

4.2. Data Preprocessing and Results of GMM Application. In this study, 1,168,307 trip records of the elderly in Seongnam during the 20 weekdays from June 10, 2019–July 5, 2019, were used. First, to filter the trip data, we obtained the card IDs of the elderly using the senior and national merit codes in the user type code. Then, only the travel records in which both the boarding and alighting area codes matched the codes of the 44 dongs in Seongnam were extracted; and there were 1,103,306 trips within Seongnam, suggesting that most trips occurred within the city. Next, to obtain the travel patterns, only the travel records of the elderly who traveled more than three days (10% of the analysis period) were extracted [39]. Finally, abnormal records with missing values or errors were eliminated. A total of 897,049 travel data went through the preprocessing stage and were used to construct a dataset for clustering.

The dataset was constructed by counting the number of boardings and alightings with respect to the dong of each user. Table 3 shows a part of the final dataset, where Card ID indicates the anonymized card id of the elderly, and the five left-hand columns denote the number of boardings at Imae, Gumi, Yatap, Seohyeon, and Sunae-dong, while the next five columns show that of alights at Seongnam, Gumi, Seohyeon, Imae, and Yatap-dong. For instance, the Card ID 3 boarded 34 times and alighted 18 times at Yatap-dong over 20 weekdays. Table 3 includes only the number of boardings and alightings for six dongs; however, the actual dataset included all 42 dongs, with a total number of 84 columns.

The t-SNE method was applied for dimension reduction before GMM clustering as the dataset exhibited a high dimensionality. As t-SNE can perfectly preserve the relative relation of data in high dimensional space, after dimensionality reduction, we can identify the patterns in low-dimensional space that exist in high-dimensional datasets [62]. The results after dimensionality reduction using t-SNE were visualized using a 2D scatter plot, and the distribution of the data can be seen in Figure 1.

Then, we applied GMM and the optimal number of clusters was derived through a sensitivity analysis based on the BIC. Figure 3 shows the results of BIC after applying GMM clustering by setting the number of clusters from 2 to 8. The smaller the BIC score, the better the model performance. Therefore, the optimal number of clusters was defined as 4, which yielded the smallest score. Figure 4 shows the results of the clustering, which categorized the elderly according to the numbers of boarding and alighting times into four clusters. Each cluster result shows which dong experienced the most boardings and alights, and Figure 3 shows that the four clusters are well classified according to the spatial travel

Table 2: ODD-based determination of autonomous shuttle service.

Classification		Available in autonomous driving	Unavailable in autonomous driving
Geometric factors	Node	Three-way intersection Four-way intersection	Over five-way intersections Roundabouts
	Number of lanes	Two-way Three-way Four-way	One-way Over five-way
	Link	Arterial road Auxiliary arterial road Collector road	Highway Underground roadway Overpass Ramp
	Road surface	Asphalt Concrete	Unpaved road Break road
Operational factors	Traffic signal	Information provision on communication networks	Unable to provide information on communication networks
	Road operation	General road section	Senior zone School zone Handicapped zone

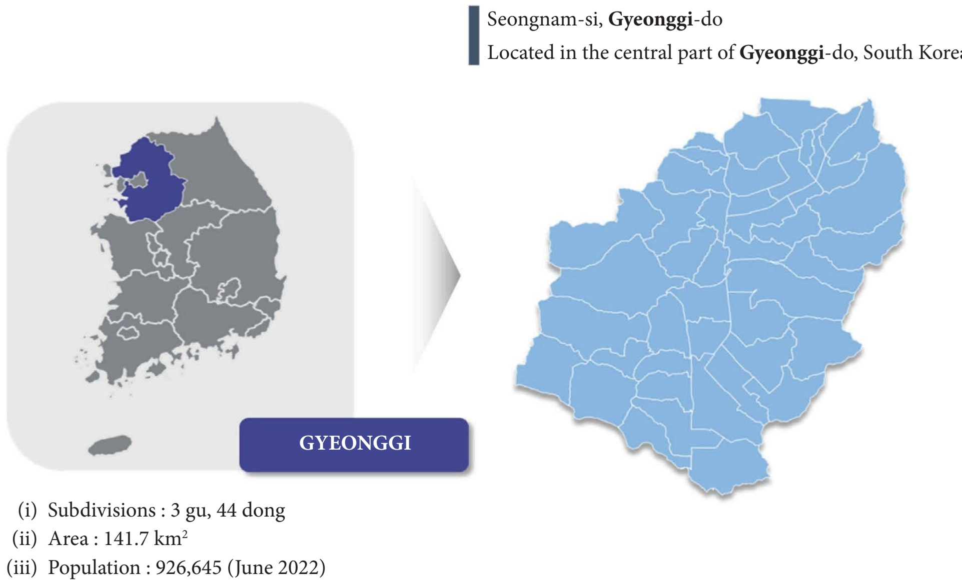


Figure 2: Seongnam description.

patterns of the elderly. In the cases of clusters 1, 2, 3, and 4, considerable boarding and alighting occurred in Gumi-dong, Yatap-dong, Jeongja-dong and its vicinity, and Seongnam-dong and Sinheung-dong, respectively.

4.3. Travel Pattern Analysis of Clusters. In the previous section, dongs with a large number of elderly boardings and alightings were identified for each cluster; then, we analyzed the spatiotemporal travel patterns for the origins and destinations with many elderly travels in each cluster. Figures 5–8 visualize the analysis results for the top ten O-D pairs according to the number of travels, and the major trip networks for each cluster was confirmed. The thickness of the link connecting each O and D denotes the number of travels, where the larger the number of journeys, the bolder the line. Figures 9–12 show the travel ratios for the top ten O-D pairs with respect to time. Based on these results, the main travel purpose for each cluster was classified.

For cluster 1, most boardings and alightings occurred in Gumi-dong, with numerous travels (55%) made between Migeum Station and Bundang Seoul National University Hospital and between Migeum Station and Bobath Hospital. In addition, travels took place between Seohyeon and Yatap Stations and the residential complex. In cluster 1, the main destination for the elderly was the hospital, with a boarding time between 09:00 to 15:00, similar to the hospital business hours; hence, the purpose of the cluster was defined as “hospital.”

The travels in cluster 2 were concentrated in Yatap-dong, more specifically, numerous travels (75%) were made between Yatap Station and the Dochon apartment complex. In addition, travels occurred between Pangyo and Moran Stations and apartment complexes, indicating that all journeys were directed to the residential areas. In addition, the travels occurred mainly in the period 16:00–19:00, marking these as return trips according to the general analysis given above. Therefore, the purpose of travel for the cluster was defined as “residence.”

For cluster 3, a large number of travels (68%) occurred between Jeongja Station and Jeongden village apartment complex, and other travels between Pangyo and Seohyeon Stations and residential complexes. In cluster 3, as in cluster 2, all travels were returning home, and the peak time was after 15:00, indicating “residence” as the purpose of travel.

In contrast to other clusters, various travel purposes were mixed in cluster 4. Travels were made between Dandae Station and Sangdaewon Market, Namhansanseong Station and Eunhaeng Market, Namhansanseong Station and the park, and Moran and Gachon University Station and the residential complex. Furthermore, the trips were concentrated in the morning and afternoon, and it appears that market visits, merchants’ commutes to and from work, park use, and return trips were combined. The trip purpose was defined as “market” along with the market travels (35%), which accounted for the largest proportion.

4.4. Implementation of Autonomous Shuttle Service. More than 94% of the travels occurred inside Seongnam, suggesting that the activity space of the elderly was within Seongnam. In addition, the average distances over which the elderly used subways and buses were 10 km and 2.6 km, respectively. Hence, in this study, a short-distance first- and last-mile autonomous shuttle route was planned. The bus and the subway travel distance distributions are shown in Figures 13 and 14, respectively.

Where multiple alighting stops were connected to one boarding stop in the O-D pair analysis, a single line was planned considering the feasibility of the short-distance operation andthe shuttle service time. The shuttle

Table 3: Part of final dataset.

Card ID	Imae-dong_O	Gumi-dong_O	Yatap-dong_O	Seohyeon-dong_O	Sunae-dong _O	Seongnam-dong_D	Gumi-dong_D	Seohyeon-dong_D	Imae-dong_D	Yatap-dong_D
1	4	19	1	1	0	1	18	2	3	1
2	1	12	4	0	0	5	12	0	1	0
3	17	0	34	0	0	17	0	0	18	18
4	2	9	3	1	1	20	0	0	0	0
5	3	6	20	0	2	33	0	4	0	13

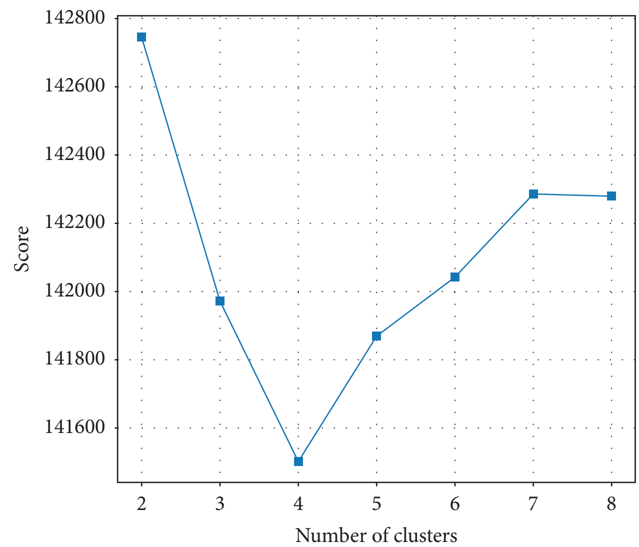


FIGURE 3: BIC scores for GMM results.

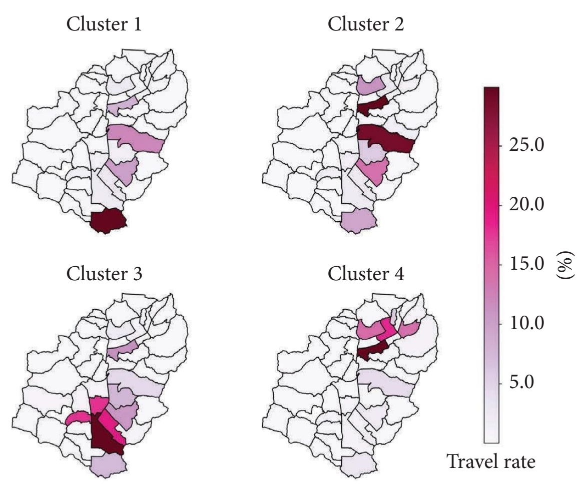


FIGURE 4: Results of GMM clustering.

operation time for each route was presented using the time section in which travels were concentrated by analyzing the travel rate between O-D pairs with respect to the time. As a large deviation arose in the travel volume between the peak and nonpeak times, time sections with a travel rate greater than or equal to the 75th percentile value were proposed. Public transportation currently operates on the proposed routes; however, if the autonomous shuttle service is introduced, this shuttle may replace or supplement several operations, depending on the service level currently provided. If the number of buses operating on the proposed route and the interval between dispatches is appropriate, some autonomous shuttles can serve as replacements to increase the operational efficiency; and if insufficient, they can supplement existing services to improve service frequency.

In addition, to demonstrate the autonomous shuttle, one major route for each travel purpose was tested according to the criteria defined in Table 2, with the major route selected to include the OD pairs with the most travel. Figures 15–18 show the travel rate with respect to time for all routes of each cluster. Figures 19–22 show the roads on which autonomous shuttles can be operated/disabled for major routes for each travel purpose.

For cluster 1, route 1 (Migeum Station-SNU Hospital), route 2 (Migeum Station-Bobath Hospital), route 3 (Seohyeon Station- Hyundai APT-Woosung Mall-Welfare Center-Miraetown APT-Beomhan Plaza), and route 4 (Yatap Station-Elementary School) can be planned. Both routes 1 and 2 start at Migeum Station; however, different routes were planned because of the difference in boarding time pattern for each line. Figure 15 shows the travel rate with respect to time for each route, and the section that exceeded the 75th percentile baseline was suggested as the shuttle service time. For example, for route 1, the shuttle service time was selected for the period 09 : 00–15 : 00, which accounts for more than 9% of the travel rate. Route 3 connects the destinations starting from Seohyeon Station,

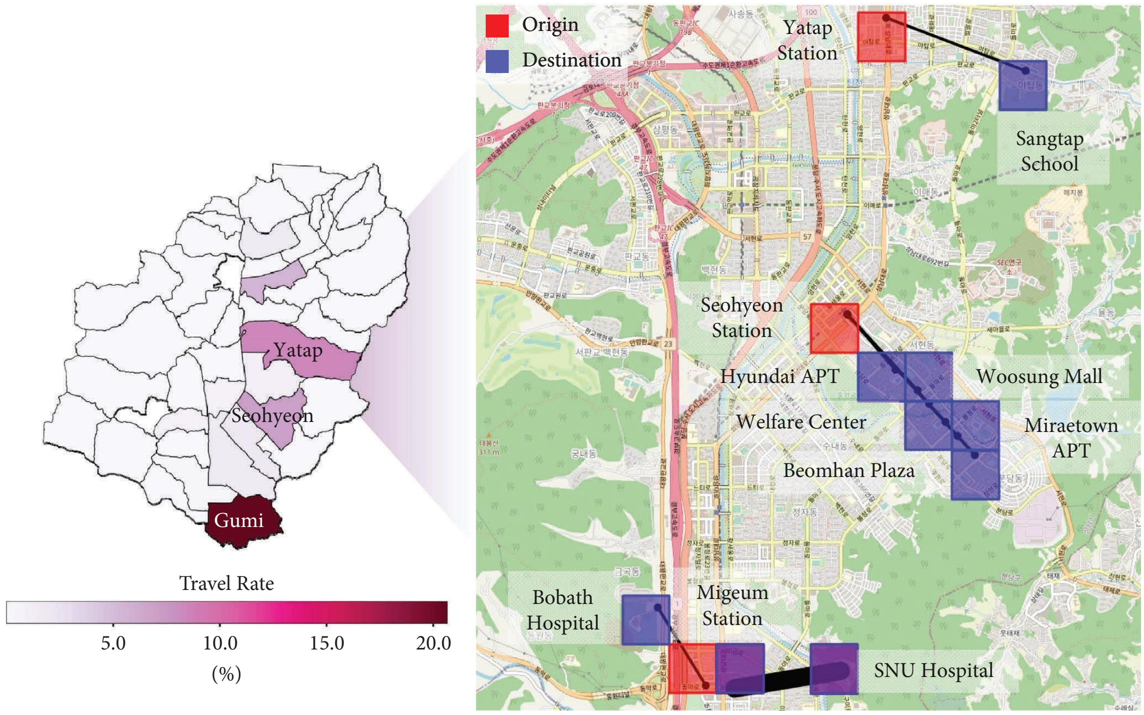


Figure 5: Top ten OD pairs of cluster 1.

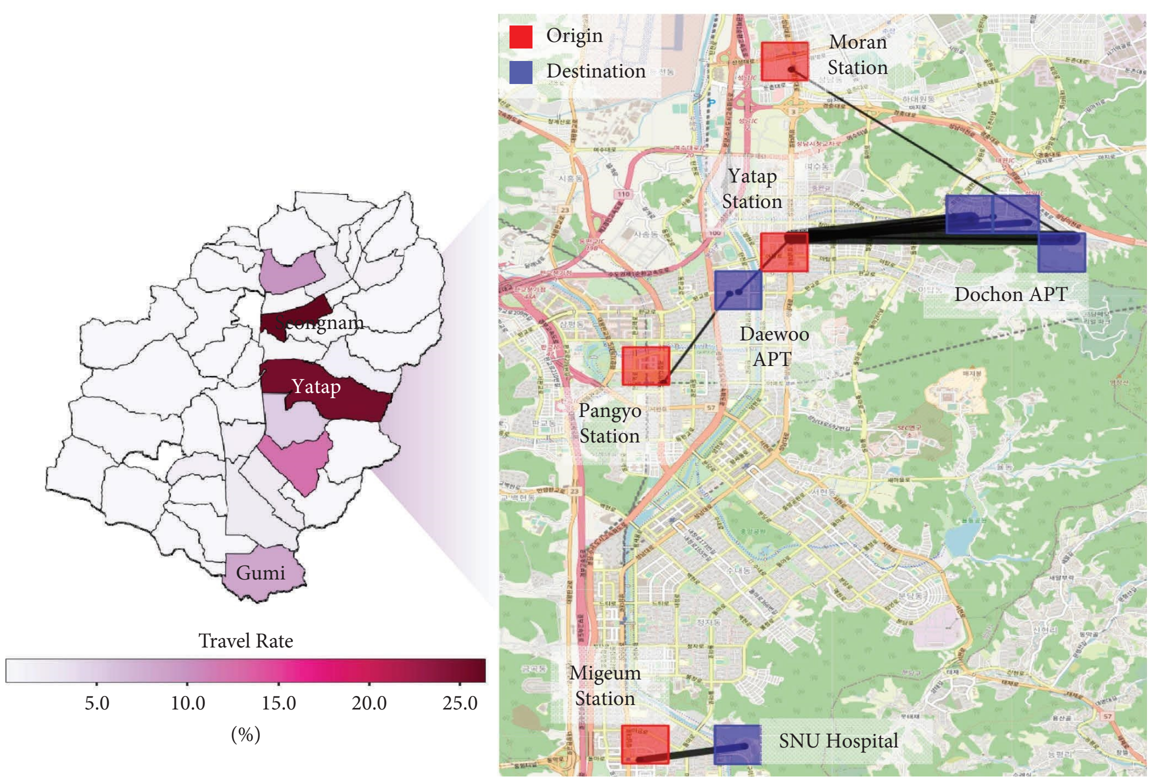


Figure 6: Top ten OD pairs of cluster 2.

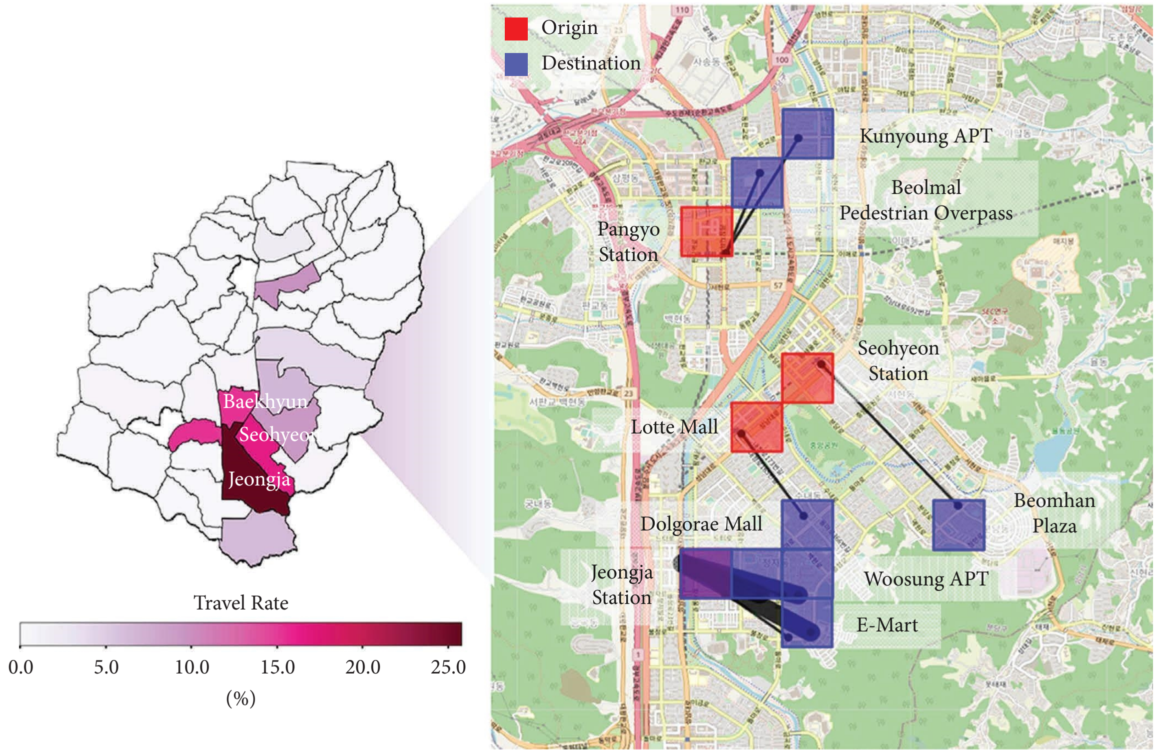


FIGURE 7: Top ten OD pairs of cluster 3.

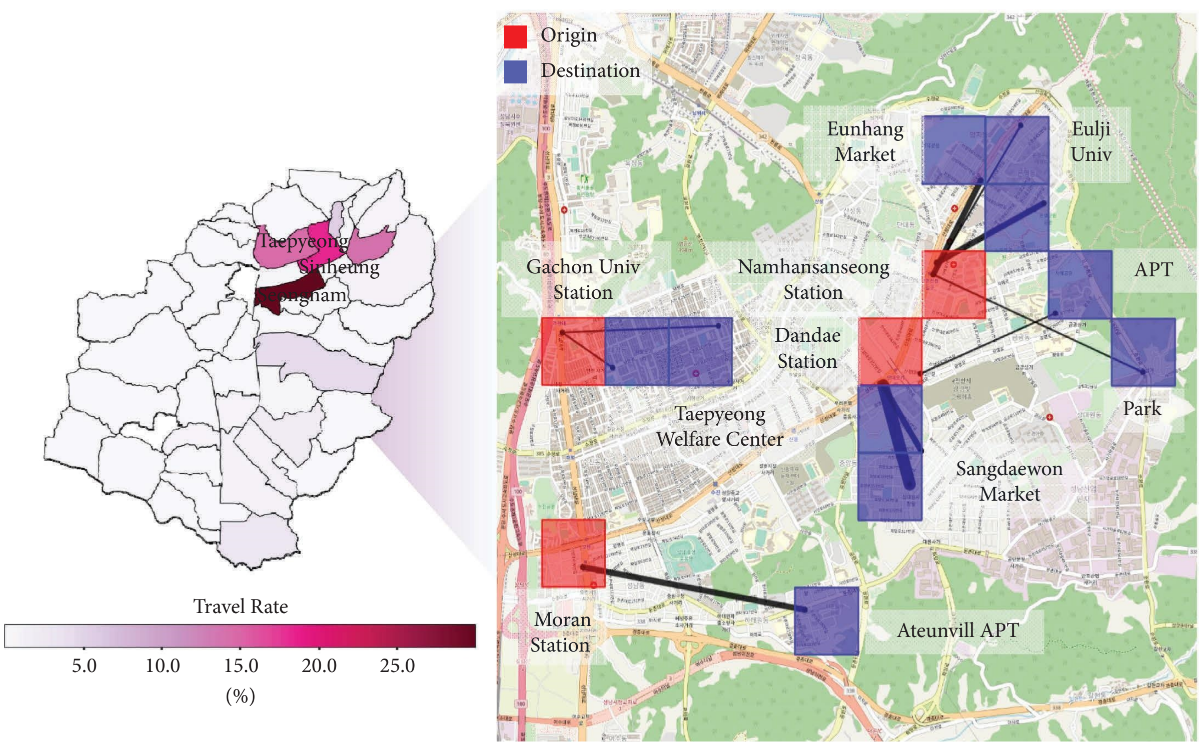


FIGURE 8: Top ten OD pairs of cluster 4.

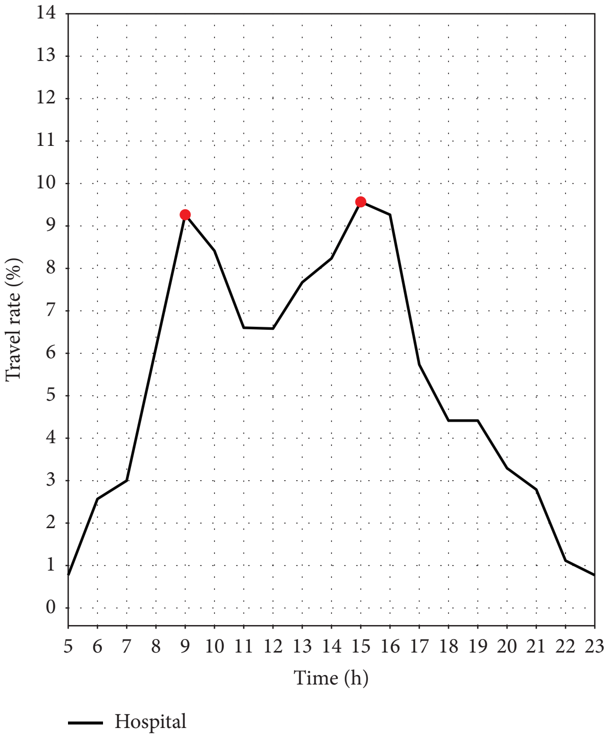


Figure 9: Hourly travel rate of cluster 1.

Figure 11: Hourly travel rate of cluster 3.

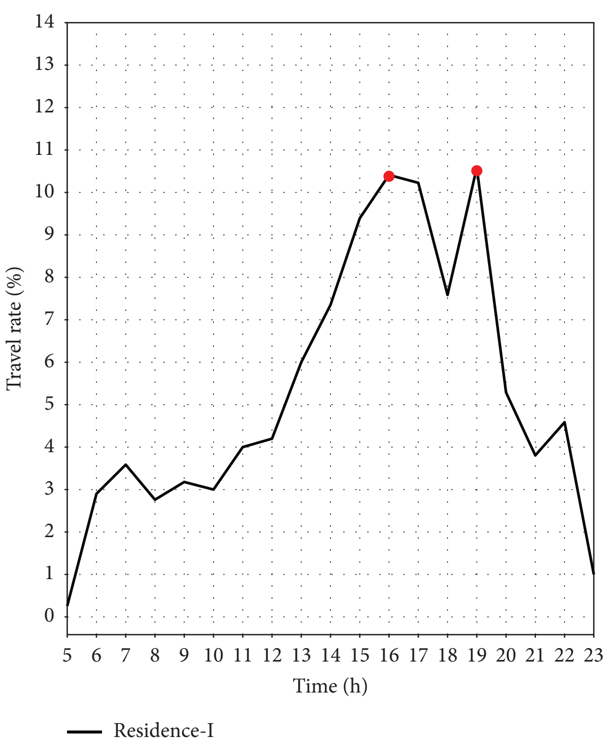


Figure 10: Hourly travel rate of cluster 2.

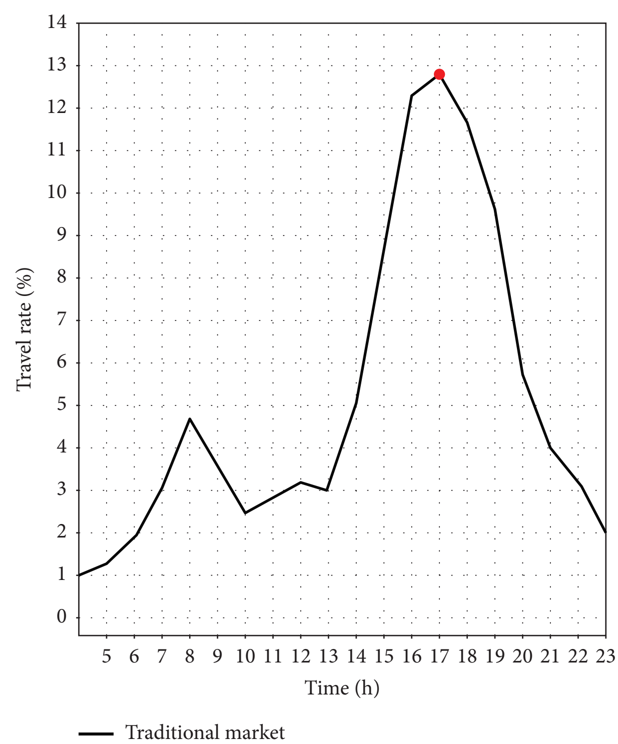


Figure 12: Hourly travel rate of cluster 4.

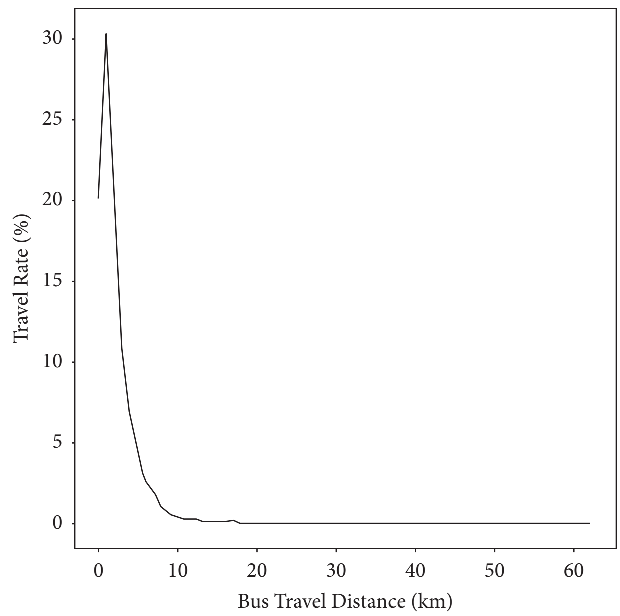


Figure 13: Bus travel distance distribution.

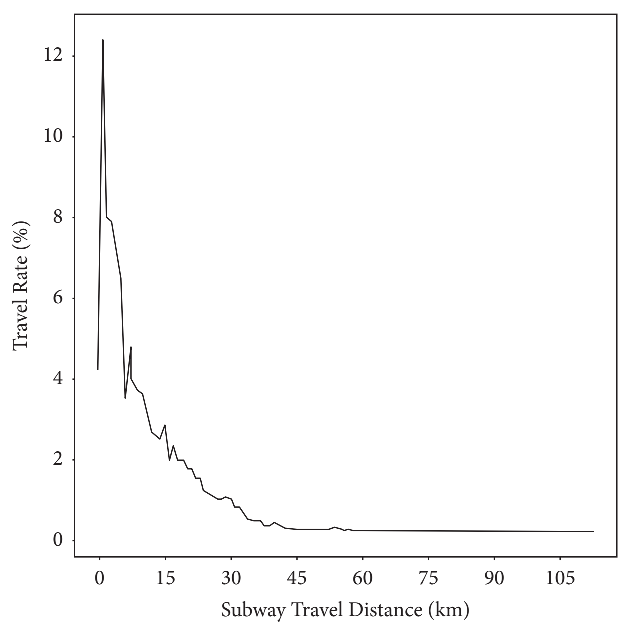


Figure 14: Subway travel distance distribution.

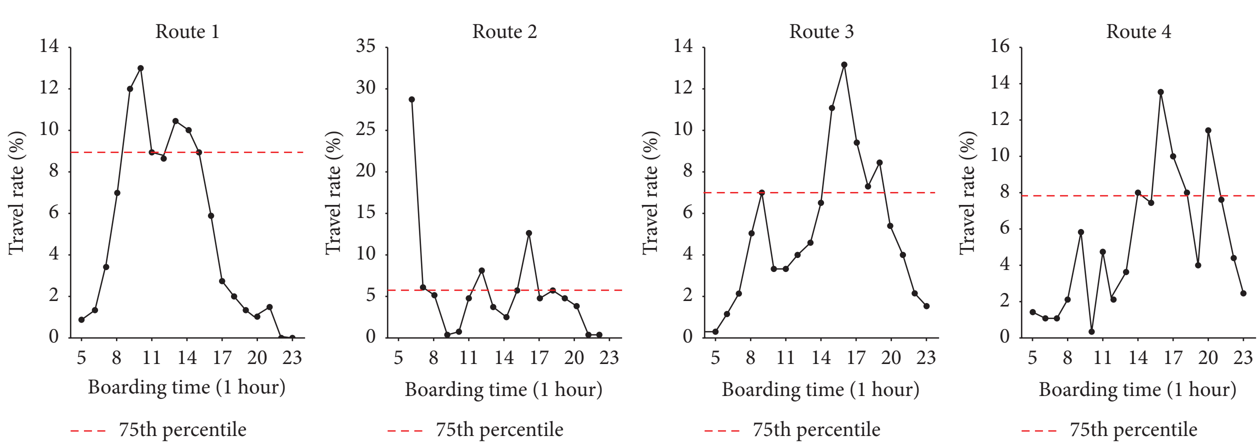


Figure 15: Travel rate by time for <hospital> routes.

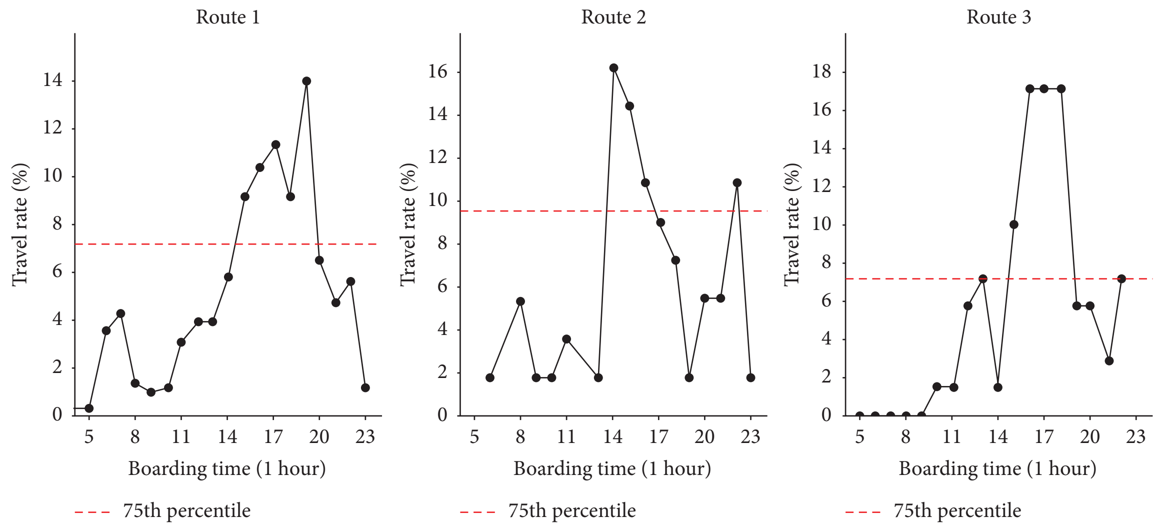


Figure 16: Travel rate by time for <residence-I> routes.

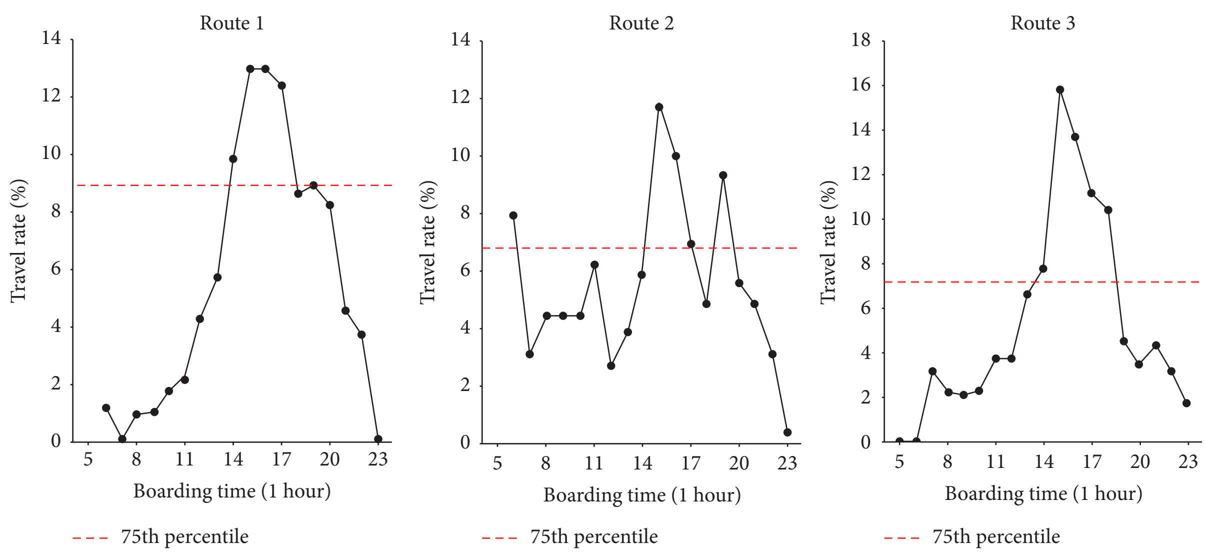


Figure 17: Travel rate by time for <residence-II> routes.

and the shuttle service time proposed was14 : 00–19 : 00. The operating time of route 4 was suggested as 14 : 00–21 : 00; and for routes 3 and 4, it appears that the characteristics of the return trips are reflected because the routes pass through a residential area. In addition, for route 2, it was judged appropriate to supplement the existing service because few bus lines are in operation. In contrast, for routes 1, 3, and 4, four or five bus lines are in operation, and the interval between dispatches is not large, so replacing certain operations with autonomous shuttles can be considered.

Figure 19 shows the results of determining whether autonomous driving is feasible using ODD components for route 1, that is, the main route for <hospital>. Ten links were classified as restricted for autonomous driving. Most of the links with school zones were excluded, and autonomous shuttles could be operated through nodes 2, 4, 6, 8, 9, and11.

For cluster 2, route 1 (Yatap Station-1,2, Danji-2 Danji-Community Center-7,8 Danji-9 Danji), route 2 (Pangyo Station-Daewoo/Keonyeong APT), and route 3 (Moran Station-Danji 9) can be planned. Route 1 connects the apartment Danjis (“apartment complexes” in Korean) in Dochon with Yatap Station and was planned as one route because most of the route travels were concentrated in the period 15 : 00–20 : 00. All three routes were for residential areas, and travels were concentrated during the afternoon owing to the nature of the return trips. For route 3, as the current service is insufficient, the autonomous shuttle is expected to complement this service.

Figure 20 shows the possible links on route 1, which passes through a residential area, with numerous one-way roads, and many sections operated as school zones. Therefore, it is necessary to switch to manual operation

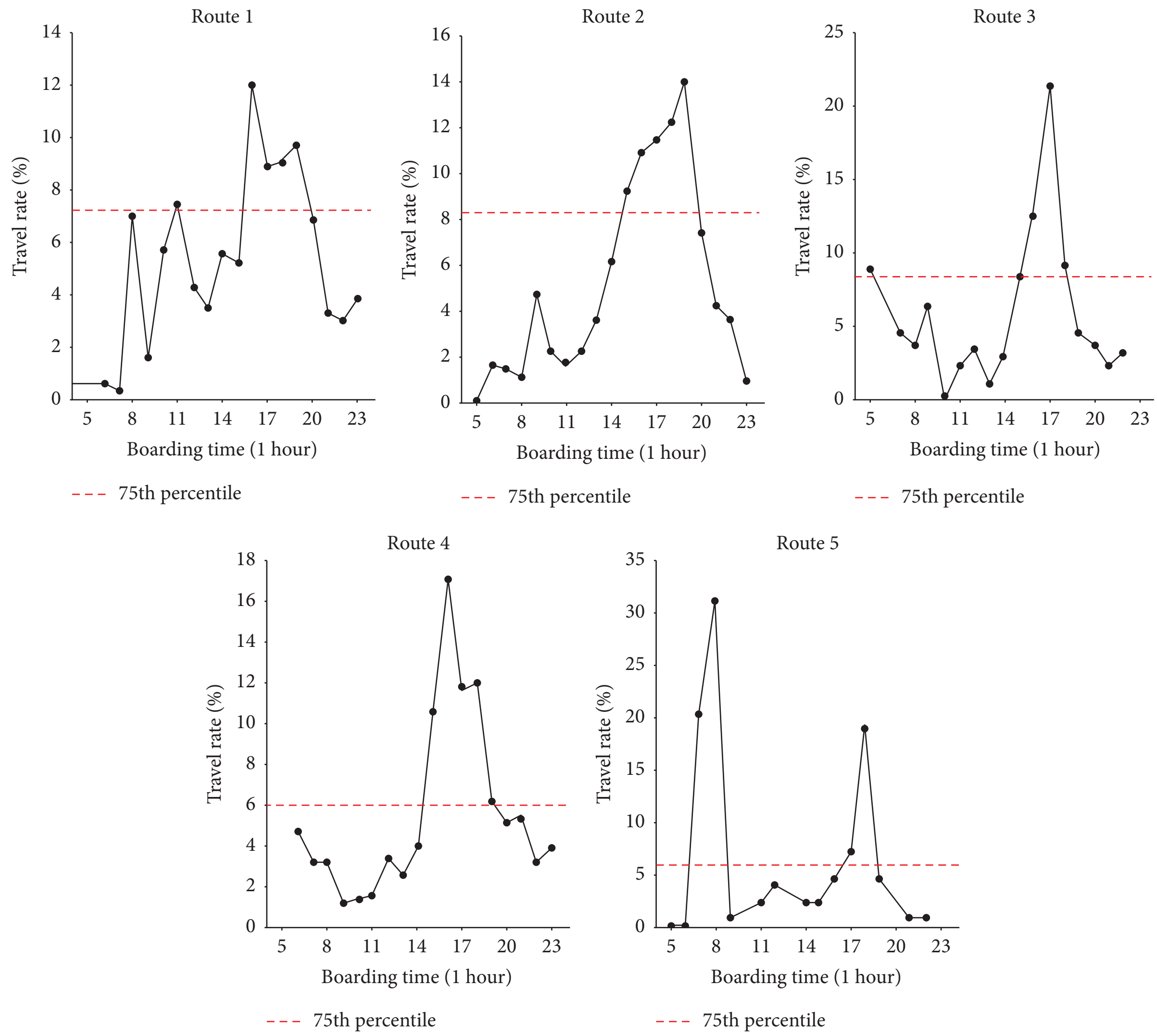
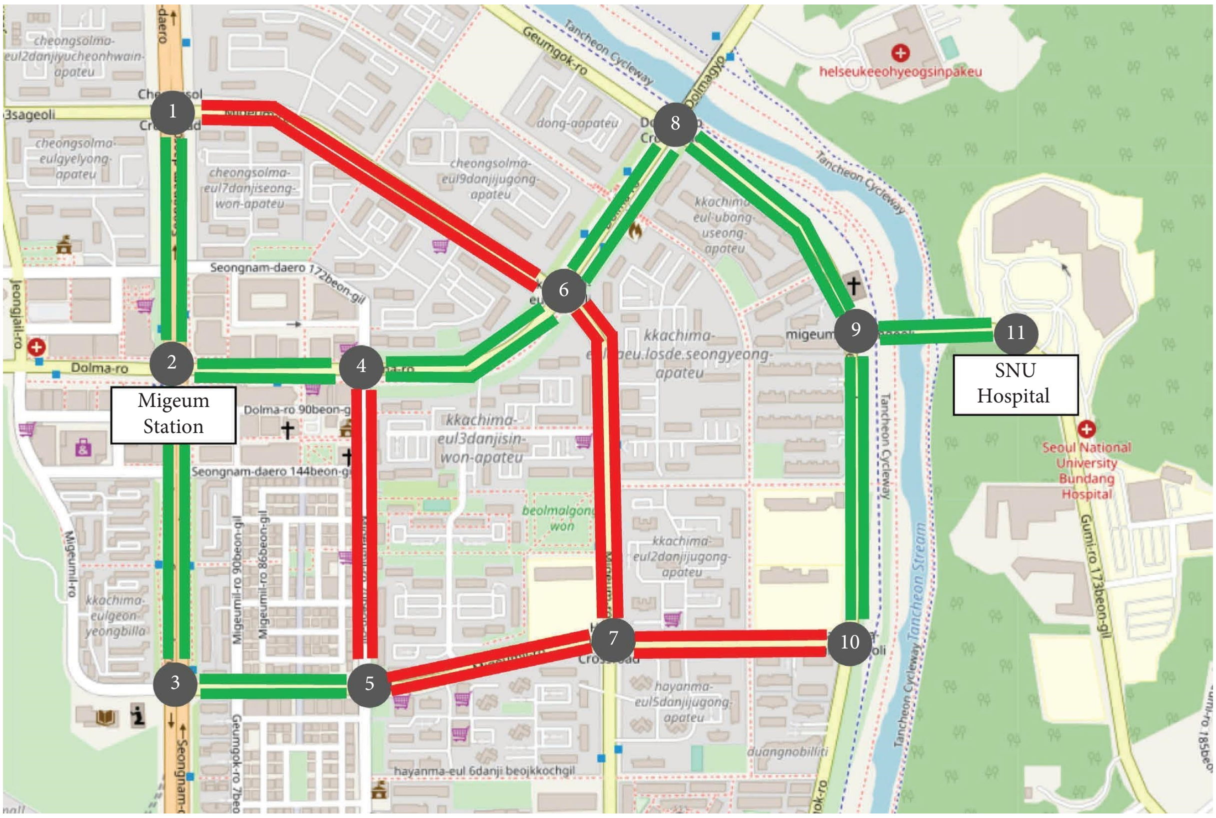


Figure 18: Travel rate by time for <traditional market> routes.

on links where autonomous shuttle operation is not possible.

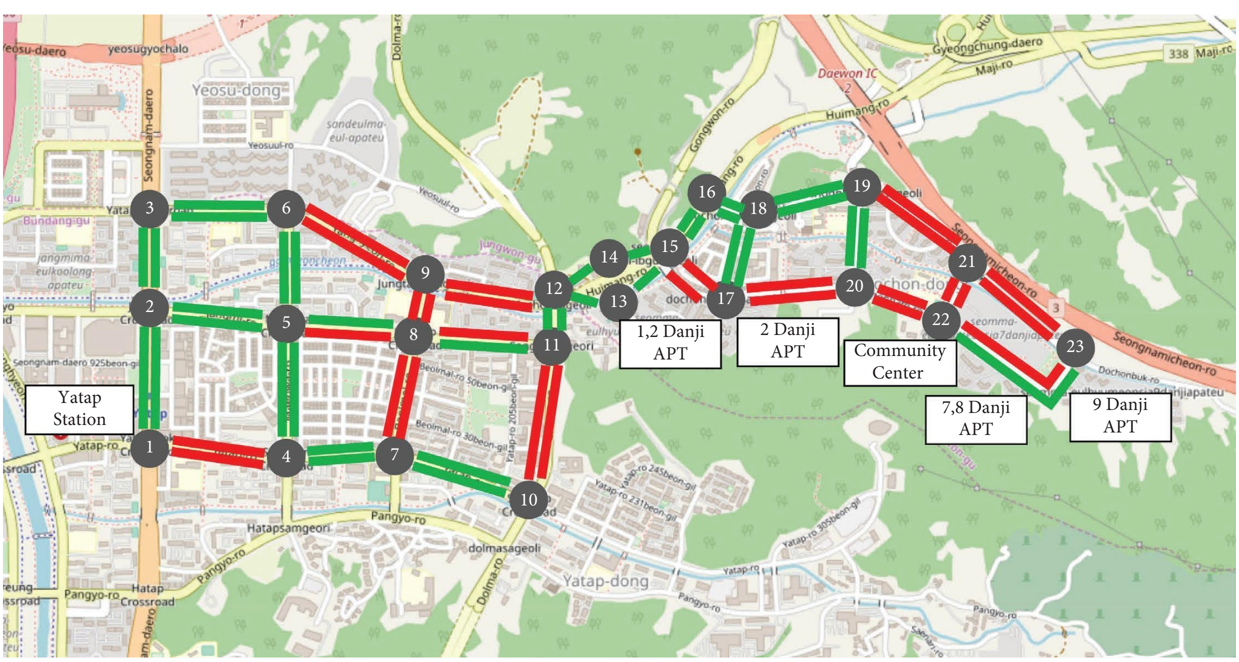
For cluster 3, route 1 (Jeongja Station-Shinhwa APT-4 Danji-6 Danji-Bundang Jungang High School-E Mart), route 2 (Pangyo Station-Beolmal Overpass-Geonyeong APT), and route 3 (Seohyeon Station-Beomhan Plaza) can be suggested. As in cluster 2, all routes head toward the residential area. As shown in Figure 17, the peak hours for all routes were 14:00–19:00, suggesting this time as the operating times of the abovementioned three routes. Furthermore, routes 1 and 2 have only two bus lines in operation. Therefore, if autonomous shuttles with high operational efficiencies are introduced on the routes, better service can be expected.

Route 1, passing the residential complexes from Jeongja Station, where the trips are most concentrated, was selected as the major route. Most links can be operated by autonomous driving; however, in the case of the link between nodes 11 and 12, it is necessary to switch to manual operation because of the school zone.



Available Node for Autonomous Shuttle
Available Link for Autonomous Shuttle
Unavailable Link for Autonomous Shuttle

FIGURE 19: Result of autonomous shuttle operation feasibility analysis based on ODD for <hospital> major route.



Available Node for Autonomous Shuttle
Available Link for Autonomous Shuttle
Unavailable Link for Autonomous Shuttle

FIGURE 20: Result of autonomous shuttle operation feasibility analysis based on ODD for <residence-I> major route.

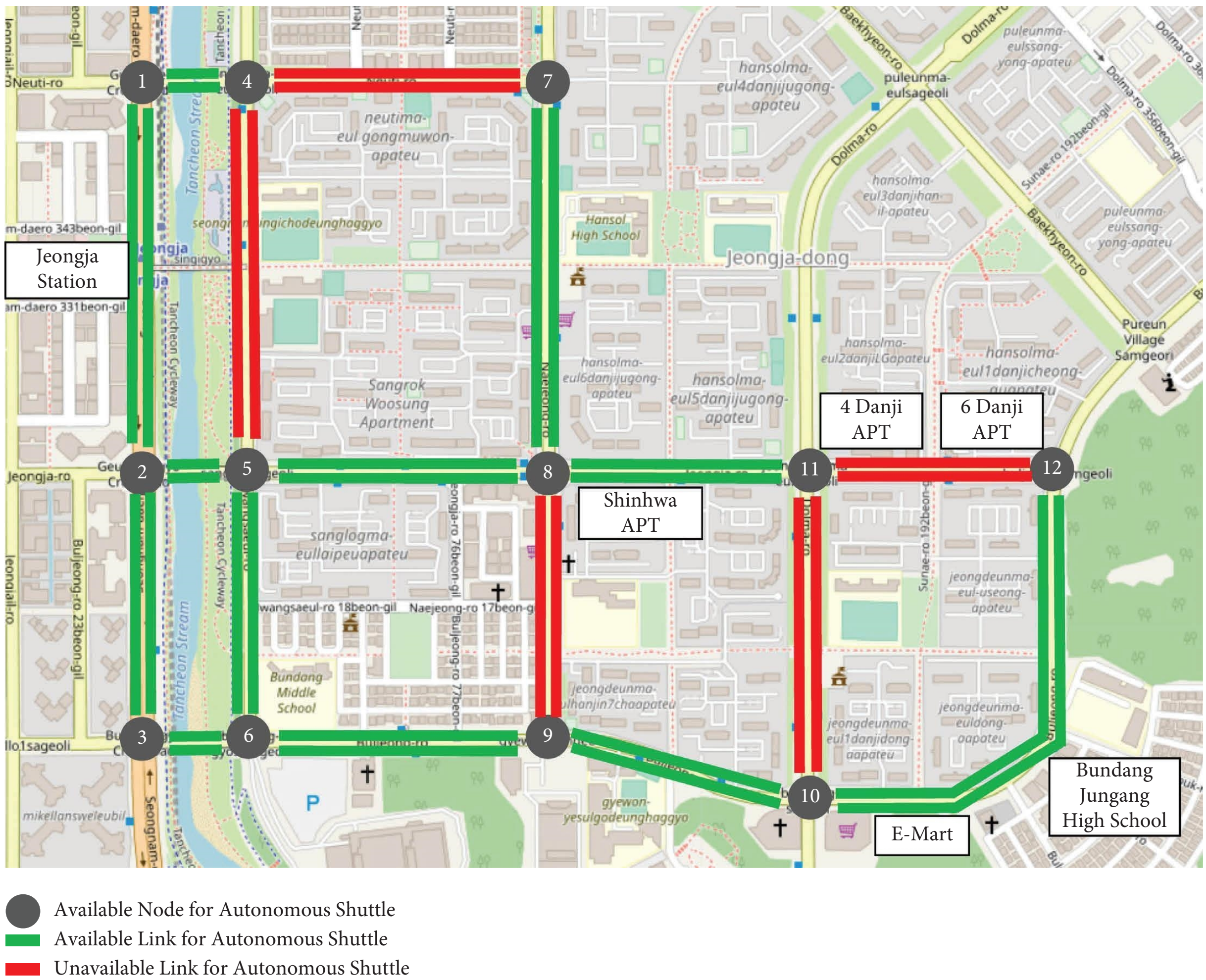


FIGURE 21: Result of autonomous shuttle operation feasibility analysis based on ODD for <residence-II> major route.

For cluster 4, the major route is route 1 (Dandae Station-Sangdaewon Market). The shuttle service time of route 1 is suggested as 16 : 00–20 : 00. Meanwhile, in addition to the major route, route 2 (Namhansanseong Station-Eunhang Market-Eulji University-Public Parking Lot), route 3 (Moran Station-Ateunvill APT), route 4 (Gacheon University Station-Taepyeong Elementary School-Welfare Center), and route 5 (Namhansanseong Station-Park) can be planned. For routes 3 and 4, which travel to the residential complex, the shuttle service times were suggested as 15 : 00–18 : 00 and 15 : 00–19 : 00, respectively. Route 5 is the only route to the park, and travels are concentrated in the morning and evening; hence, operating hours of 07 : 00–09 : 00 and 17 : 00–19 : 00 were suggested. Most of the suggested routes provide sufficient public transportation services. However, for route 4, supplementation is necessary because of the large dispatch interval.

The result of determining whether autonomous driving is possible on route 1 is shown in Figure 22. Links included in the senior and school zones and links with one-way lanes (links 8-9 and 11-9) were excluded, making autonomous shuttle operation impossible.

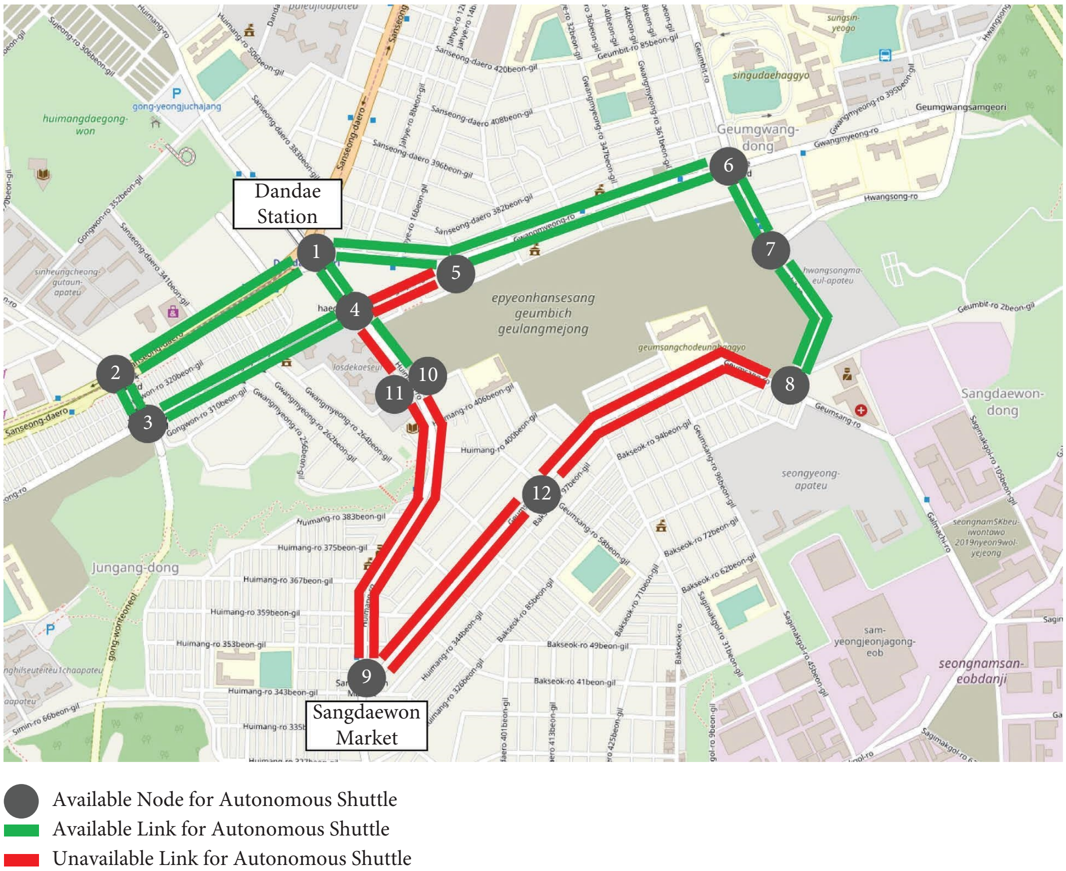


Figure 22: Result of autonomous shuttle operation feasibility analysis based on ODD for <traditional market> major route.

5. Conclusions

The transportation-disadvantaged population is rapidly increasing because of aging, and the mobility of the elderly is being enhanced by life expectancy improvements and lifestyle changes. To respond to these demands, in this paper, we proposed a comprehensive framework to introduce autonomous customized shuttles for the elderly based on their travel pattern. In particular, the first- and the last-mile shuttle service was proposed to incorporate the demands of the elderly who mainly travel short distances and to increase the accessibility of the existing transportation methods. We conducted a case study on Seongnam, where the mobility services for the elderly are insufficient. Using the smart card data, the elderly were clustered according to their travel patterns, and autonomous shuttle routes and service times were suggested for each travel purpose.

The dataset consisting of the boarding and alighting patterns of the elderly was clustered using GMM and we derived four travel types through a sensitivity analysis based on the BIC. By analyzing the spatiotemporal travel patterns of the O-D pairs for each cluster, the travel purposes were classified as hospital, residence-I, residence-II, and traditional market. In the case of <hospital>, many travels (55%) between Migeum Station and the hospital occurred, and the travel rate was high during hospital business hours. In <residence-I>, most of the travels (75%) occurred between Yatap Station and Dochon APT Danjis in the late afternoon. For <residence-II>, all travels, including those between the residential complexes and Jeongja Station (68%), corresponded to the return trips. In the <traditional market>, travels from the station to the traditional market were predominant (35%), and the travels were concentrated in the early morning and late afternoon.

O-D pair analysis revealed that most boarding stops were near the subway station. Thus, we planned the routes by connecting the boarding stop near the subway station to the alighting stops. The possible routes were presented for each cluster, and for each major route, we analyzed the road links where autonomous shuttles can be operated using ODD. It was deemed necessary to switch to manual operations for some routes passing through residential areas because of the one-way lanes passing through school and senior zones.

We believe that this study provides useful insights for introducing autonomous customized mobility services for the travel patterns of the elderly. Moreover, it is expected that the methodology can be applied to other areas in Gyeonggi-do, where the mobility support services for the elderly are vulnerable, by comparing the results to provide shuttle services.

However, owing to the characteristics of the smart card data, the travel records of elderly people only taking the bus or using cash only were not identified, so the proposed routes may be limited. Currently, the cash usage rate of city buses in the metropolitan area, including Seoul, is low at about 1% and the smart card data can be considered to represent almost the entire passengers. But still, there are some elderly people who use cash only and their travel behaviors cannot be reflected. Therefore, we plan to utilize additional travel records that can be collected regardless of the payment method or transport modes such as household travel surveys and floating population data. Furthermore, in future research studies, we intend to optimize our methodology for designing routes suitable for actual operation by minimizing the user and operator costs based on the genetic algorithms.

Data Availability

The data used to support the findings of the study can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Institute for Information & Communication Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (Grant no. 2021001415).

References

[1] Ministry of Land, *Infrastructure and Transport, Survey of Moving Convenience for Transport Vulnerables*, Ministry of Land, Battaramulla, Sri Lanka, 2021.

[2] L. Han, W. Zhou, and W. Li, "City as a major source area of fine particulate (PM2.5) in China," *Environmental Pollution*, vol. 206, no. 2, pp. 183–187, 2015.

[3] The Korea Transport Institute, *Better Driving License Management and Mobility Services for the Elderly*, The Korea Transport Institute, Sejong City, Korea, 2020.

[4] Seoul Metropolitan Council, *Seoul Metropolitan Council Budget and Policy Magazine*, Seoul Metropolitan Council, Seoul, South Korea, 2021.

[5] F. Kovacs, S. McLeod, and C. Curtis, "Aged mobility in the era of transportation disruption: will autonomous vehicles address impediments to the mobility of ageing populations?" *Travel, Behaviour and Society*, vol. 20, pp. 122–132, 2020.

[6] R. Zandieh and R. A. Acheampong, "Mobility and healthy ageing in the city: exploring opportunities and challenges of autonomous vehicles for older adults' outdoor mobility," *Cities*, vol. 112, Article ID 103135, 2021.

[7] Gyeonggi Research Institute, *Financial Support Standards for Gyeonggi-Do Customized Bus*, Gyeonggi Research Institute, Gyeonggi-do, South Korea, 2019.

[8] Y. Lyu, C. Chow, V. Lee, J. Ng, Y. Li, and J. Zeng, "CB-planner: a bus line planning framework for customized bus systems," *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 233–253, 2019.

[9] Z. Sun, K. Zhou, X. Yang, X. Peng, and R. Song, "Optimization method of customized shuttle bus lines under random condition," *Algorithms*, vol. 14, no. 2, 2021.

[10] T. Kusakabe and Y. Asakura, "Behavioural data mining of transit smart card data: a data fusion approach," *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 179–191, 2014.

[11] J. Zhao, Q. Qu, F. Zhang, C. Xu, and S. Liu, "Spatio-temporal analysis of passenger travel patterns in massive smart card data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 11, pp. 3135–3146, 2017.

[12] H. Nishiuchi, Y. Kobayashi, T. Todoroki, and T. Kawasaki, "Impact analysis of reductions in tram services in rural areas in Japan using smart card data," *Public Transport*, vol. 10, no. 2, pp. 291–309, 2018.

[13] M. Eltved, N. Breyer, J. Ingvardson, and O. Nielsen, "Impacts of long-term service disruptions on passenger travel behaviour: a smart card analysis from the Greater Copenhagen area," *Transportation Research Part C: Emerging Technologies*, vol. 131, Article ID 103198, 2021.

[14] J. Huang, D. Levinson, J. Wang, J. Zhou, and Z. J. Wang, "Tracking job and housing dynamics with smartcard data," *Proceedings of the National Academy of Sciences*, vol. 115, no. 50, pp. 12710–12715, 2018.

[15] H. Faroqi, M. Mesbah, J. Kim, and A. Tavassoli, "A model for measuring activity similarity between public transit passengers using smart card data," *Travel Behaviour and Society*, vol. 13, pp. 11–25, 2018.

[16] M. K. El Mahrsi, E. Côme, L. Oukhellou, and M. Verleysen, "Clustering smart card data for urban mobility analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, pp. 712–728, 2017.

[17] F. Devillaine, M. Munizaga, and M. Trépanier, "Detection of activities of public transport users by analyzing smart card data," *Transportation Research Record*, vol. 2276, no. 1, pp. 48–55, 2012.

[18] X. Ma, Y. Wu, Y. Wang, F. Chen, and J. Liu, "Mining smart card data for transit riders' travel patterns," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 1–12, 2013.

[19] S. Lee and G. Lee, "Development and application of the flow cluster detection method: a case of public transit trips in Seoul," *Journal of the Korean Cartographic Association*, vol. 21, no. 2, pp. 129–144, 2021.

[20] S. Liu, T. Yamamoto, E. Yao, and T. Nakamura, "Examining public transport usage by older adults with smart card data: a longitudinal study in Japan," *Journal of Transport Geography*, vol. 93, Article ID 103046, 2021.

[21] F. Shao, Y. Sui, X. Yu, and R. Sun, "Spatio-temporal travel patterns of elderly people – a comparative study based on buses usage in Qingdao, China," *Journal of Transport Geography*, vol. 76, pp. 178–190, 2019.

[22] T. Yoon, Y. Lim, J. Park, H. Lee, K. Yeon, and K. Jung, "Strategies to implement autonomous public transit: focusing on autonomous bus rapid transit and shuttle service," *Korea Research Institute for Human Settlements*, vol. 19, 2019.

[23] S. Lim, W. Jang, S. Tak, and T. Park, *Strategies of Public Transport Service in the Era of Autonomous Vehicle and Shared Transport*, The Korea Transport Institute, Sejong City, Korea, 2018.

[24] K. Gurumurthy and K. Kockelman, "Analyzing the dynamic ride-sharing potential for shared autonomous vehicle fleets using cellphone data from Orlando, Florida," *Computers Environment and Urban Systems*, vol. 71, pp. 177–185, 2018.

[25] F. Dandl, B. Grueber, H. Friese, and K. Bogenberger, "Design and simulation of a public-transportation-complimentary

autonomous commuter shuttle," *Transportation Research Procedia*, vol. 41, pp. 240–250, 2019.

[26] K. Rehrl and C. Zankl, "Digibus: results from the first self-driving shuttle trial on a public road in Austria," *European Transport Research Review*, vol. 10, no. 2, pp. 51–11, 2018.

[27] P. S. Lin, A. Kourtellis, N. Menon, C. Chen, and R. Rangaswamy, *Campus Automated Shuttle Service Deployment Initiative (No. NCTR Livability PPPR# 18)*, University of South Florida, Tampa, FL, USA, 2020.

[28] S. Haustein, "Mobility behavior of the elderly: an attitude-based segmentation approach for a heterogeneous target group," *Transportation*, vol. 39, no. 6, pp. 1079–1103, 2012.

[29] R. Alsnih and D. A. Hensher, "The mobility and accessibility expectations of seniors in an aging population," *Transportation Research Part A: Policy and Practice*, vol. 37, no. 10, pp. 903–916, 2003.

[30] A. Siren and L. Hakamies-Blomqvist, "Private car as the grand equaliser? Demographic factors in Finnish men and women aged 65?" *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 7, no. 2, pp. 107–118, 2004.

[31] M. Lee, J. Han, and H. Lee, "Analysis of the transit ridership pattern using transportation card data: focusing on ganghwa," *The Journal of The Korea Institute of Intelligent Transport Systems*, vol. 17, no. 2, pp. 58–72, 2018.

[32] B. Melit Devassy and S. George, "Dimensionality reduction and visualisation of hyperspectral ink data using t-SNE," *Forensic Science International*, vol. 311, Article ID 110194, 2020.

[33] C. Yi, S. Tuo, S. Tu, and W. Zhang, "Improved fuzzy C-means clustering algorithm based on t-SNE for terahertz spectral recognition," *Infrared Physics & Technology*, vol. 117, Article ID 103856, 2021.

[34] S. K. Popat, "Review and comparative study of clustering techniques," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 1, pp. 805–812, 2014.

[35] S. Saraswathi and L. M. I. Sheela, "A comparative study of various clustering algorithms in data mining," *International Journal of Computer Science and Mobile Computing*, vol. 11, no. 11, pp. 422–428, 2014.

[36] S. Aghabozorgi, A. Seyed Shirkhorshidi, and T. Ying Wah, "Time-series clustering – a decade review," *Information Systems*, vol. 53, pp. 16–38, 2015.

[37] X. Zhou, H. Zhang, G. Ji, and G. Tang, "A multi-density clustering algorithm based on similarity for dataset with density variation," *IEEE Access*, vol. 7, pp. 186004–186016, 2019.

[38] S. Ben Salem, S. Naouali, and Z. Chtourou, "A fast and effective partitional clustering algorithm for large categorical datasets using a k-means based approach," *Computers & Electrical Engineering*, vol. 68, pp. 463–483, 2018.

[39] I. Lee, J. Min, K. Kim, and S. Kho, "Generating travel patterns of public transportation users using a k-means clustering based on smart card data," *Journal of the Korean Society for Railway*, vol. 23, no. 3, pp. 204–215, 2020.

[40] I. Lee, S. Cho, K. Kim, S. Kho, and D. Kim, "Travel pattern-based bus trip origin-destination estimation using smart card data," *PLoS One*, vol. 17, no. 6, Article ID 270346, 2022.

[41] P. N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Pearson: Addison Wesley, Boston, MA, USA, 2006.

[42] Z. Wang, C. Da Cunha, M. Ritou, and B. Furet, "Comparison of K-means and GMM methods for contextual clustering in HSM," *Procedia Manufacturing*, vol. 28, pp. 154–159, 2019.

[43] S. Majumder, "A Gaussian mixture model method for eigenvalue-based spectrum sensing with uncalibrated multiple antennas," *Signal Processing*, vol. 192, Article ID 108404, 2022.

[44] A. Mirzal, "Statistical analysis of microarray data clustering using NMF, spectral clustering, kmeans, and GMM," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 2, pp. 1173–1192, 2022.

[45] E. Patel and D. Kushwaha, "Clustering cloud workloads: K-means vs. Gaussian mixture model," *Procedia Computer Science*, vol. 171, pp. 158–167, 2020.

[46] B. Greve, I. Pigeot, I. Huybrechts, V. Pala, and C. Börnhorst, "A comparison of heuristic and model-based clustering methods for dietary pattern analysis," *Public Health Nutrition*, vol. 19, no. 2, pp. 255–264, 2016.

[47] A. Briand, E. Côme, M. Trépanier, and L. Oukhellou, "Analyzing year-to-year changes in public transport passenger behaviour using smart card data," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 274–289, 2017.

[48] Y. Ji, Y. Cao, Y. Liu, W. Guo, and L. Gao, "Research on classification and influencing factors of metro commuting patterns by combining smart card data and household travel survey data," *IET Intelligent Transport Systems*, vol. 13, no. 10, pp. 1525–1532, 2019.

[49] M. Lin, Z. Huang, T. Zhao, Y. Zhang, and H. Wei, "Spatiotemporal evolution of travel pattern using smart card data," *Sustainability*, vol. 14, no. 15, p. 9564, 2022.

[50] F. Fabozzi, S. Focardi, S. Rachev, B. Arshanapalli, and M. Hoechstoetter, "The basics of financial econometrics," *The Frank J. Fabozzi Series*, John Wiley and Sons, New York, NY, USA, 1st edition, 2014.

[51] Y. Khan, P. Khan, S. Kumar, J. Singh, and R. M. Hegde, "Detection and spread prediction of COVID-19 from chest X-ray images using convolutional neural network-Gaussian mixture model," in *Proceedings of the IEEE 17th India Council International Conference (INDICON)*, pp. 1–6, New Delhi, India, December 2020.

[52] T. Chen and J. Zhang, "On-line multivariate statistical monitoring of batch processes using Gaussian mixture model," *Computers and Chemical Engineering*, vol. 34, no. 4, pp. 500–507, 2010.

[53] H. Jung and S. Lee, "Analysis of traffic characteristics for the transportation vulnerable," *Journal of The Korean Society of Civil Engineers*, vol. 33, no. 1, pp. 241–249, 2013.

[54] S. Lee and M. Hickman, "Trip purpose inference using automated fare collection data," *Public Transport*, vol. 6, no. 1-2, pp. 1–20, 2013.

[55] A. Alsger, A. Tavassoli, M. Mesbah, L. Ferreira, and M. Hickman, "Public transport trip purpose inference using smart card fare data," *Transportation Research Part C: Emerging Technologies*, vol. 87, pp. 123–137, 2018.

[56] K. S. Shehadeh, H. Wang, and P. Zhang, "Fleet sizing and allocation for on-demand last-mile transportation systems," *Transportation Research Part C: Emerging Technologies*, vol. 132, Article ID 103387, 2021.

[57] N. Quarles, K. Kockelman, and M. Mohamed, "Costs and benefits of electrifying and automating bus transit fleets," *Sustainability*, vol. 12, no. 10, p. 3977, 2020.

[58] P. Camps-Aragó, L. Temmerman, W. Vanobberghen, and S. Delaere, "Encouraging the sustainable adoption of autonomous vehicles for public transport in Belgium: citizen acceptance, business models, and policy aspects," *Sustainability*, vol. 14, no. 2, p. 921, 2022.

[59] E. Thorn, S. C. Kimmel, M. Chaka, and B. A. Hamilton, *Framework for Automated Driving System Testable Cases and*

Scenarios, National Highway Traffic Safety Administration, Washington, DC, USA, 2018.

[60] H. Kim, K. Lim, J. Kim, and W. Son, "Operational design domain for testing of autonomous shuttle on arterial road," *The Journal of The Korea Institute of Intelligent Transport Systems*, vol. 19, no. 2, pp. 135–148, 2020.

[61] Gyeonggi Research Institute, *Plan to Enhance the Transportation Convenience of Mobility Disadvantaged Persons*, Gyeonggi Research Institute, Gyeonggi-do, South Korea, 2020.

[62] J. Tang and X. Yan, "Neural network modeling relationship between inputs and state mapping plane obtained by FDA–t-SNE for visual industrial process monitoring," *Applied Soft Computing*, vol. 60, pp. 577–590, 2017.