Driving Risk Assessment Using Non-Negative Matrix Factorization With Driving Behavior Records

Hyunwoo Seo, *Student Member, IEEE*, Jongkyung Shin, *Student Member, IEEE*, Ki-Hun Kim, Chiehyeon Lim, *Member, IEEE*, and Jungcheol Bae

***Abstract*— Aggressive driving behavior (ADB) is a major cause of traffic accidents. As ADB is controllable, ADB-based driving risk assessment is an effective method for drivers and transportation companies to ensure driving safety. Conventionally, the relationships between ADBs and accident-related records are analyzed when assessing driving risk. However, such records typically overlook driver responsibility for driving risks and depend considerably on the person producing the data (e.g., police officers or insurance managers). Foremost, conventional approaches do not consider non-accident situations that comprise most driving scenarios. Thus, we propose a novel driving risk assessment method that uses only ADB data. In this method, interpretable latent risk factors are extracted from ADB data via sparse non-negative matrix factorization (NMF), and then the driving risk score is computed on a scale of 0–100. The proposed method was validated by adopting a real-world application to assess the driving risk of bus drivers in South Korea and by conducting an evaluation performed by transportation experts in conjunction with the Korea Transportation Safety Authority. Results revealed that the proposed method can discriminate between high- and low-risk driving, thus providing clear guidelines to improve driving. Then, the proposed driving risk score assessment method using NMF was compared with existing machine learning-based risk assessment methods. The proposed method outperformed the conventional methods in terms of driving risk discrimination and interpretability. This study can provide risk assessment guidelines based on driving behavior records and contribute to the application of machine learning in transportation safety management.**

Manuscript received 24 March 2021; revised 13 October 2021 and 29 March 2022; accepted 22 June 2022. Date of publication 1 August 2022; date of current version 7 November 2022. This work was supported in part by the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT under Grant 2020R1C1C1014312 and Grant 2020R1F1A1076067; by the Ministry of Education under Grant 2020S1A5A8046717, Grant 2021R1I1A4A01049121, and Grant 2022S1A5A8054010; and in part by the Institute of Information & Communications Technology Planning & Evaluation funded by the Korean government under Grant 2020-0-01336 (Artificial Intelligence Graduate School Program - UNIST). The Associate Editor for this article was L. M. Bergasa. *(Corresponding authors: Chiehyeon Lim; Ki-Hun Kim.)*

Hyunwoo Seo is with the Department of Industrial Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, Republic of Korea.

Jongkyung Shin is with the Graduate School of Artificial Intelligence, Ulsan National Institute of Science and Technology, Ulsan 44919, Republic of Korea.

Ki-Hun Kim is with the Department of Industrial Engineering, Pusan National University, Busan 43241, Republic of Korea (e-mail: kihun@pusan.ac.kr).

Chiehyeon Lim is with the Department of Industrial Engineering and the Graduate School of Artificial Intelligence, Ulsan National Institute of Science and Technology, Ulsan 44919, Republic of Korea (e-mail: chlim@unist.ac.kr).

Jungcheol Bae is with the Korea Transportation Safety Authority, Seoul 03937, Republic of Korea.

This article has supplementary downloadable material available at https://doi.org/10.1109/TITS.2022.3193125, provided by the authors.

Digital Object Identifier 10.1109/TITS.2022.3193125

***Index Terms*— Driving risk assessment, aggressive driving behavior, driving behavior record, non-negative matrix factorization.**

I. Introduction

TRAFFIC accidents impede traffic flow and cause urban problems worldwide. The U.S. National Highway Traffic Safety Administration reports that more than 5,000,000 car crashes occur in the United States annually, with approximately 30% of accidents leading to serious injuries or fatalities [1]. Most traffic accidents are caused by inappropriate driving behaviors and driver errors [2], [3]. In particular, aggressive driving behavior (ADB) is a major cause of driving accidents [4] (Fig. 1). Studies have shown that the ADB management of drivers can effectively enhance traffic safety [5], [6]. For example, ADB-based driving risk assessment enables individual drivers to reassess their risky driving behaviors. Moreover, as ADB is controllable, ADB-based driving risk assessment enables drivers to manage and improve their driving behaviors. Transportation companies can use this approach when evaluating employee driving safety and providing specialized training for specific driving behaviors. The aforementioned merits indicate the importance of ADB-based driving risk assessment [7].

In previous research, driving risk was assessed by analyzing the relationships between ADBs and accident-related records (e.g., [8]–[10]). Fig. 1 shows ADBs and accident-related records that have been treated as independent and dependent variables, respectively, in existing studies. However, the use of dependent variables is unreliable and inaccurate in driving risk assessment because accident records are subjective and they depend considerably on the person generating the data (e.g., police officers or insurance managers). Furthermore, accident-related records are archived securely but separately by government and private organizations, which renders their acquisition difficult. Non-accident situations that comprise most driving scenarios should also be considered in driving risk assessment. For example, crash occurrence records have been used for driving risk assessment (e.g., [8], [11]). However, this approach can lead to overestimation or underestimation. Crashes occur irrespective of safe driving. Crash occurrence records are not automatically recorded when crashes occur and are unlikely to contain sufficient information concerning all occurrences; therefore, these records tend to be under-reported and contain biases [12]. The limitations of crash data and crash-based methods suggest the necessity of

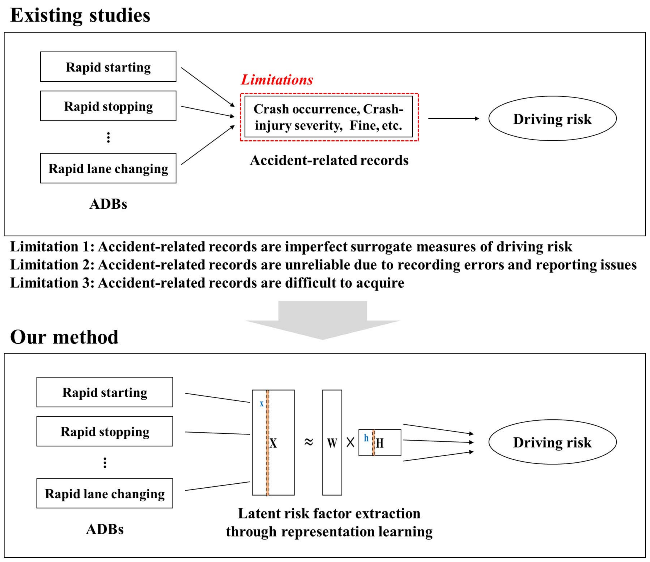


Fig. 1. Driving risk assessment using ADB measurement.

a driving risk assessment method that is not predicated on crash occurrences or other accident-related records, including the severity of injury and amount of fines [13].

In this study, a novel driving risk assessment method was proposed for gauging driving risk on a trip (i.e., from the beginning to the end of driving) through a quantitative score by using only ADB measurements. Specifically, this method aggregates ADB measurements into a driving risk score using non-negative matrix factorization (NMF), which is a representation learning method for extracting interpretable latent factors from a non-negative dataset via matrix factorization [14]. The application of NMF to a dataset of ADB measurements obtained from multiple trips enables the identification of latent factors that can explain patterns of risky driving behavior. In the proposed approach, the frequencies of ADBs are measured, and each latent factor identified via NMF is represented by the non-negative weighted sum of these ADB frequencies. The driving risk score of a trip is computed by aggregating the values of the latent risk factors. Non-negativity implies a positive relationship among the frequencies of each ADB, its latent factor value, and the corresponding driving risk score, thereby enhancing the interpretability of the score and its practical utility. In view of improving the discrimination and interpretability of the driving risk score, specific constraints are imposed in the NMF process to transform the original ADB measurements into latent risk factor values.

To the best of our knowledge, this study is the first one to assess driving risk based on a score by using only driving behavior records. Therefore, the proposed method can assess driving risk without utilizing accident-related records, which are typically inaccurate or difficult to acquire. The methodological contribution of the proposed method was validated through comparative experiments with existing risk assessment methods developed in other domains. The sparse NMF used by the proposed method is superior in terms of discriminating and interpreting driving risk scores, which contributes to incorporating machine learning in accident research and transportation safety management. This method can also be applied to risk assessment in other domains. The contribution of our work was further validated through its real-world application in conjunction with Korea Transportation Safety Authority (KOTSA). We also confirmed the necessity, validity, and applicability of the proposed approach through expert evaluation. By using the proposed method, drivers can easily monitor ADB occurrences that increase their driving risk scores. The drivers can effectively lower their scores by reducing such ADB occurrences. Furthermore, the proposed method can be used by transportation companies to manage the driving safety of their employees, thus enhancing the overall traffic safety. The development of a reliable and practical method for driving risk assessment has been a significant challenge that KOTSA wishes to address.

The remainder of this paper is organized as follows. Previous related work is reviewed in Section II. Section III explains the proposed method. Section IV describes the real-world application of the proposed method. Section V presents an evaluation of the method via comparative experiments. Section VI describes the expert surveys and interviews and analyzes the outcomes. Section VII discusses the implementation issues related to the proposed method. Section VIII provides the conclusion of the study.

II. Related Work

First, Section II-A describes the existing studies on ADBs and driving risk assessment. Then, Section II-B explains risk scoring methods based on representation learning. Finally, Section II-C introduces the basic elements of NMF as a preliminary information for Section III.

A. ADBs and Driving Risk Assessment

ADB is defined as a deliberate behavior motivated by impatience, annoyance, and hostility or an attempt to save time, which is likely to increase collision risk [15]. Many existing studies have identified ADB in various settings and proven that ABD is a major driving risk [16], [17]. Shinar and Compton [18] performed site investigations and examined the characteristics of drivers in relation to the frequencies of ADBs. Abou-Zeid *et al.* [19] identified the factors influencing ADBs by using a driving simulator to generate certain events in the traffic environment and evaluated the reactions of drivers to those events. Čabarkapa *et al.* [20] performed a questionnaire survey and revealed that ADB is a predictor of traffic accidents based on hierarchical regression analysis. However, the analyses of driving data collected from artificial settings provide driving behavior measures that differ from those in actual driving environments, which result in biased outcomes. Furthermore, ADB occurrences measured in contrived settings may not be applicable to on-road situations; thus, ADBs in the actual driving environments must be investigated [15], [21].

The development of in-vehicle and mobile phone sensors has facilitated the real-time and continuous collection of driving behavior data in driving environments [22], [23]. For example, by using in-vehicle sensors installed in commercial vehicles in South Korea, ADB occurrences can be measured in real time to characterize the driving environment. KOTSA performed dynamics analysis and validation of ADBs on actual roads and defined the ten most significant ones and their measurement criteria for trucks, buses, and taxis [24]. ADBs can be measured using driving records obtained from digital tachographs (DTGs). For example, for commercial buses, short-term over-speeding is defined as driving at a speed 20 km/h greater than the speed limit. These ADBs can be measured using DTG data, such as revolutions per minute (RPM) and global positioning system (GPS) records (see Supplementary Material B for further details on ADBs and the use of DTGs in South Korea).

Considerable research has been conducted on driving risk assessment based on ADB measurements in driving environments. For example, Osafune *et al.* [11] proposed the use of ADBs as driving risk indices to statistically differentiate safe and risky drivers. Safe and risky drivers were classified according to the number of accidents in the preceding five years and their driving experience. Islam and Mannering [8] focused on crash-injury severity as a driving risk and modeled it using random-parameter multinomial logit models. They utilized three years of real-world crash data and defined crash–injury severity as a target variable in three levels, namely, no injury, minor injury, and severe injury. They also investigated the differences in crash-injury severity in accordance with aggressive and nonaggressive driving behaviors. Guo and Fang [25] considered a naturalistic driving experiment setting in which subjects drive in actual driving environments to evaluate risk factors using the crash and near-crash (CNC) frequency, and then they classified the overall risk of individual drivers based on their CNC rates. Wang *et al.* [10] defined driving risk as a potential threat that can cause vehicle crashes and other accidents and classified driving risk as high risk, moderate risk, and low risk. A near-crash database was developed based on the naturalistic driving experimental data. Chen *et al.* [26] proposed a non-negativity-constrained autoencoder to predict driving risk on trips. A total of 76 driving behavior time series were considered for each trip, and their non-negativity-constrained autoencoder learned representative and distinct hidden driving behavior features, allowing for the classification of the driving risk level of a trip.

The aforementioned studies have analyzed historical crash records as surrogate measures to evaluate driving risk. However, these records are limited in terms of accurately quantifying driving risk. First, the driving risk of crash-free trips, which constitute most cases in driving records, can not be accurately assessed. Second, crash occurrence records are not perfect measures of driving risk. In particular, a crash may occur when individuals drive safely, and it may not occur when people drive rashly. As crash records are not automatically recorded when crashes occur, they are unlikely to contain accurate information concerning all crashes [12]. Furthermore, the person responsible for recording an incident may incorrectly report the driver who is at fault or even under-report the crashes [27]. Although near-crashes can be used in a naturalistic driving experiment setting as auxiliary surrogate measures to crash events [28], near-crash data cannot be quantitatively obtained in most cases [29]. Furthermore, selecting crashes and near-crashes without introducing bias is difficult [30]. Therefore, driving risk assessment that is independent of crash records is essential (see Supplementary Material A for further details on the limitations of using accident-related records).

Several studies based on unsupervised learning have been recently developed in the domain of driving risk. Mantouka *et al.* [31] adopted the two-stage clustering approach to obtain unsafe driving profiles. Li *et al.* [32] extracted driving patterns from driving sequence data by using unsupervised Bayesian algorithms and clustering. Hossain *et al.* [33] applied association rule mining (ARM) and joint corresponding analysis to investigate the behavioral patterns of teen drivers involved in crashes. Although the aforementioned studies were able to identify unsafe driving patterns, they did not assess each trip by using a quantitative driving risk score. Hossain *et al.* [33] specified the crash severity level in the crash database as the consequent in ARM for the analysis. Thus, it is not a fully unsupervised learning approach.

In summary, driving risk is conventionally assessed by analyzing incomplete, unreliable, or inaccurate surrogate measures of driving risk, while the unsupervised approaches in driving risk research mostly investigate driving patterns. In this study, a novel driving risk assessment method using only ADB measurements with unsupervised learning is proposed. As mentioned earlier, ADBs comprise the major factors for determining driving risk levels [4]; they constitute behavioral variables that can be controlled by drivers and subsequently used for practical driving safety management [34]. Using ADBs alone in driving risk assessment enables us to focus on the responsibility of drivers themselves for determining risk. A previous study defined driving risk as the summation of the products of the intensities and frequencies of ADBs [7]. However, if this definition is followed, then the various effects of each ADB on driving risk and the dependencies among ADBs cannot be ascertained. Multiple ADBs frequently and simultaneously occur, and they influence driving risk through their interactions [27]. Thus, driving risk should be assessed using latent factors that optimally represent the ADB dataset.

B. Representation Learning for Risk Assessment

Obtaining reliable and accurate measures of risk, denoted as dependent (Y) variables, is challenging in various domains such as healthcare and finance. Given this difficulty, numerous studies have focused on using representation learning to assess the risk of a given subject by using only independent (X) variables [35]. The common scoring methods include principal component analysis (PCA) and autoencoders. PCA identifies new features that preserve as much as possible the variance of an original dataset and that are orthogonal to one another [36]. Autoencoders, a type of artificial neural network, include an encoder–decoder architecture and determine a latent representation (encoding) to closely reconstruct their input (decoding) [37].

Anderloni *et al.* [38] developed a household financial vulnerability scoring system based on nonlinear PCA. Two principal components were identified, and they could explain approximately 70% of the variance in a dataset collected from a survey on household financial distress. In their work, the financial risk score was defined as the sum of the values of the two principal components. Choi *et al.* [39] constructed an aggregate air quality score that ranked states by their levels of air pollutants. After applying PCA to the data of five air pollutants in U.S. states, the score was defined as the normalized value of the first principal component, which explained more than 80% of the variance. Jia *et al.* [40] and Zhang *et al.* [41] defined biological age (i.e., health risk) based on biomarkers of aging about the aging process of vital organs as the first principal component with an eigenvalue greater than 1. Li *et al.* [35] constructed a composite sustainability indicator for manufacturing companies by using principal components with cumulative variance greater than 90%. Despite the success of the aforementioned studies in evaluating the risk of given subjects, they lacked an interpretation of how the risk can be addressed; that is, the principal components were not semantically interpreted, and risk management strategies were not proposed. This limitation can be attributed to the PCA mechanism; although principal components are produced to explain the largest variance within a dataset and to be orthogonal to each other, the variance may be semantically vague [42], [43]. Thus, interpretation of each component is difficult, and the selection of a small number of components causes information loss from the dataset.

With advancements in deep learning, risk scores have also been developed using various forms of autoencoders. Nguyen *et al.* [44] constructed abnormality scores for human walking gaits by using a sparse deep autoencoder. Three autoencoders were used to model the X, Y, and Z axes of a gait dataset. The score was defined as a weighted sum of reconstruction errors attributable to the three autoencoders. This approach is reasonable, especially since the input of an abnormal walking gait should result in a greater reconstruction error than the input of a normal walking gait. Nguyen and Meunier [45] further improved their study by using an adversarial autoencoder. The human walking abnormality score was defined as the weighted sum of the reconstruction loss, the probability that a sample was extracted from a prior distribution, and the discriminator output. Xu *et al.* [46] developed a health indicator of a rotating machinery by using the output of a stacked autoencoder to reconstruct vibration signals. Then, a health indicator was constructed using the reconstructed output. In the aforementioned studies, autoencoders of various architectures were used to satisfy the input characteristics and score the definitions, but they employed the common approach of transforming original features into latent factors through a complex combination of nonlinear activation functions. This complexity hinders the interpretation of the relationship between the original features and latent factors. Besides, when latent factors are aggregated into a score, the complex relationship further complicates the understanding of the original features that predominantly influence the score.

NMF is an effective tool for addressing the limitations of PCA and autoencoders and therefore was used in this study as a tool for representation learning in driving risk assessment. NMF identifies latent factors that can explain all parts of a dataset as the non-negative weighted sum of the original features [42], [47]. Furthermore, NMF can convert the ADB dataset into driving risk scores to represent non-negative combinations of latent factors in an intuitive and interpretable manner.

C. NMF

NMF is a matrix factorization method designed for analyzing data matrices whose elements are non-negative [42], [48]. Given a non-negative data matrix $X \in \mathbb{R}^{N\times M}$ of N features and M samples, NMF decomposes X into the product of a basis matrix $W \in \mathbb{R}^{N\times K}$ and an encoding matrix $H \in \mathbb{R}^{K\times M}$ as follows:

$$X \approx WH \tag{1}$$

where K represents the number of latent factors and is smaller than N or M. Note that both W and H are non-negative. Eq. 1 can be expressed as $x \approx Wh$, where x and h are columns of X and H, respectively. Thus, each data vector x is represented by a linear combination of columns of W, and, therefore, W can be interpreted as a basis matrix for the linear approximation of the data in X. Each column in H is the latent representation with respect to the new basis W.

The non-negative constraints on W and H permit only additive combinations of multiple bases. Consequently, NMF can learn a parts-based representation compatible with the intuitive notion of combining parts to form a whole. This property of parts-based representations is useful in many real-world applications including document clustering, face analysis, and recommender systems [49]–[51]. Various NMF loss functions have been proposed to obtain optimal representations; among these approaches the following squared error function is the most frequently used approach:

$$L(W,H) = ||X - WH||^2 = \sum_{i,j}(X_{ij} - (WH)_{ij})^2 \tag{2}$$

This loss function is convex on either W or H but not on both variables, rendering to obtain a globally optimal solution difficult. However, many optimization methods have been proposed to determine the local minima. Paatero *et al.* [52] proposed a gradient algorithm to minimize the loss function. Lee and Seung [42] devised a multiplicative algorithm for the original NMF that is simple to implement and can ensure excellent performance. The multiplicative update rule for the squared error function is defined as follows:

$$\begin{aligned} W_{ik} &\leftarrow W_{ik}\frac{(XH^T)_{ik}}{(WHH^T)_{ik}}, \\ H_{kj} &\leftarrow H_{kj}\frac{(W^TX)_{kj}}{(W^TWH)_{kj}} \end{aligned} \tag{3}$$

In this update rule, $L(W, H)$ is nonincreasing, and its convergence has been proven by [48].

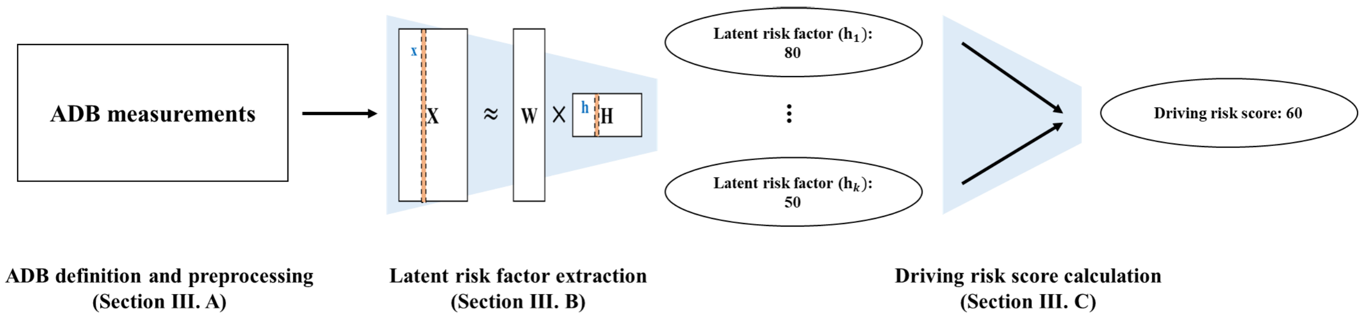


Fig. 2. Driving risk assessment using NMF.

III. Driving Risk Assessment Using NMF

The proposed method of driving risk assessment by using NMF involves a sequential procedure from ADB measurement to risk score calculation (Fig. 2). Specifically, the procedure consists of the following steps: (i) ADB definition and preprocessing, (ii) latent risk factor extraction, and (iii) driving risk score calculation. These three steps are explained in detail in Sections III-A, III-B, and III-C.

A. ADB Definition and Preprocessing

In assessing human behavior, accurate and reliable descriptions of behavior variables should be defined initially [53], [54]. Without clearly defining human behavior variables, inaccurate and uninterpretable assessment is inevitable. Therefore, the integrity of independent behavior variables is necessary in behavior analysis. In the case of driving, various ADB criteria should be defined according to the vehicle type and driving context [24]. Based on the ADB definitions, the frequencies of ADBs are extracted from the in-vehicle driving records for each trip. In-vehicle sensors generally operate from the time the engine is started to the time it is stopped, and this span is recognized as a trip. Thus, the frequency of an ADB X_{ADB} per trip is measured as follows:

$$X_{ADB} = \frac{total\ duration\ of\ an\ ADB\ (s)}{duration\ of\ the\ trip\ (h)} \tag{4}$$

According to [55], drivers typically do not perform ADBs. Therefore, ADB frequency distributions are generally right-skewed. While most trips have low ADB frequencies, a few others have high ADB frequencies. These high-frequency trips can be identified as outliers. Furthermore, noisy driving records containing irrelevant information, such as refueling, can include ADBs with frequencies that slightly deviate from those of normal driving records. Therefore, preprocessing is required to identify and remove trips with slightly deviating and high-frequency ADBs. Before driving risk score calculation, a variety of techniques, such as outlier detection, noise removal, or data transformation should be applied depending on the circumstances to obtain accurate ADB measurement.

B. Latent Risk Factor Extraction

The extraction of latent risk factors from ADB measurements is critical in determining the optimal combinations of ADBs that represent the ADB dataset and consider the various effects of ADBs on driving risk. In this study, we used sparse NMF to factorize the ADB dataset matrix X of N ADBs and M trips into the product of a basis matrix W and an encoding matrix H. Then, the *L1*-norm regularization term was applied to the original NMF to impose sparsity on W, as follows:

$$L(W, H) = ||X - WH||^2 + \lambda||W||_1 \tag{5}$$

subject to the constraints $\forall ikj : W_{ik} \geq 0, H_{kj} \geq 0$ and $||H_{:,j}|| = 1$, where $H_{:,j}$ denotes the j-th column of H, the regularization parameter $\lambda \geq 0$, W_{ik} is the influence of ADB i on the latent risk factor k, and H_{kj} is the value of latent risk factor k of trip j. Sparse NMF is useful for dimensionality reduction, feature extraction, and source separation. The sparsity constraints on W can reduce small loadings and induce large loadings, which decreases the *L1*-norm regularization term. This phenomenon produces latent risk factors that are predominantly represented by major ADBs, thus considerably improving the interpretability of the latent risk factors. Although the original NMF yields multiple equivalent factorizations that lead to various interpretations [49], the unit norm constraint on each column of H fixes the scale of H to enable the sparse NMF to have a unique factorization. The formulation proposed in this study is analogous to the one utilized by Hoyer [56], and the proofs of the optimization procedure and its convergence can be found in the said reference.

A large value of K reduces the loss (i.e., reconstruction error), but it results in a complex NMF mechanism and a factor extraction method that cannot compress the ADB dataset. By contrast, a small value of K generally increases the loss, but it simplifies the model, which lead to factor extraction method that cannot accurately represent the ADB dataset. Thus, the optimal value of K should be selected while considering the complexity and validity of the NMF model. In determining K, a rank selection method based on the minimum description length (MDL) [57] is considered to be appropriate. MDL is a method for selecting between models of varying complexities based on information theory. The aim of MDL is to determine a simple model that can retain the information of the data as much as possible, and its objective is to select the optimal K as the rank to minimize the message length. Thus, the optimal K is the number of latent risk factors that ensure both the interpretability and validity of the method. MDL fits well with the requirement of driving risk assessment under which a trade-off exists between the error and model size. In this case, the error and model size represent the validity and complexity of the model, respectively.

NMF is the most suitable method for learning latent factors for driving risk score calculation without using accident-related records. ADB frequency is typically low because most drivers do not perform ADBs frequently, which leads to a sparse ADB dataset. The driving risk score should satisfy the intrinsic relationship in which a higher ADB frequency corresponds to a higher driving risk [4], [58]. However, dense representations of latent factors that allow negative entries may not consider the inherent sparsity of the ADB dataset and the relationship between ADB frequency and driving risk. The basis and encodings of NMF entail a large fraction of vanishing coefficients due to the non-negativity constraints forcing the coefficient of irrelevant features to become zero [42]. In addition, non-negative values in the basis and encodings of NMF can identify positive relationships among ADB frequency, latent risk factors, and risk score, all of which are useful for the interpretation and real-world application of learning results. Furthermore, NMF does not make statistical assumptions other than the non-negativity, unlike other feature extraction methods such as PCA in which the orthogonality between latent factors is assumed. Such statistical assumptions can destroy the inherent properties of representations of real-world driving patterns. Finally, NMF does not require multiple user-defined hyperparameters except the number of latent variables, which makes it easy to use for driving risk score assessment in different contexts (e.g., different locations, times, and drivers) with high reproducibility.

C. Driving Risk Score Calculation

Following the NMF mechanism, $x_j = Wh_j = \sum_{k=1}^{K} w_k h_{kj}$. The ADB of each trip is represented as a linear combination of latent risk factors. The values in the columns in the encoding matrix H are the weights of the latent risk factors in a trip and indicate the scores of the driving risk components. Therefore, the average of the values in the columns of H can represent the driving risk on a trip evaluated in terms of the frequency of ADBs during the trip. Thus, a raw driving risk score can be constructed for each trip i as follows:

$$Raw\ driving\ risk\ score\ (RDR^i) = \frac{1}{K}\sum_{k=1}^{K}\hat{h_{ki}} \tag{6}$$

where $\hat{h_{ki}}$ is the standardized value of trip i for the kth risk factor. With standardization, $\hat{h_{ki}}$ represents the relative score of each driving risk component. The raw score is normalized from 0 to 100 for convenient interpretation, as follows:

$$Driving\ risk\ score\ (DR^i) = \frac{RDR^i - min(RDR)}{max(RDR) - min(RDR)} \tag{7}$$

where 0 and 100 denote the safest and riskiest driving, respectively. The driving risk score is defined as the relative driving risk on a trip relative to other trips evaluated in terms of ADB frequency.

IV. Real-World Application

We collaborated with KOTSA to apply the proposed method to assess the driving risk on bus trips in regions of South Korea. Public transit buses must be safe because they are used by a large number of passengers. However, traffic accidents involving buses occur frequently in South Korea. Overall, 874.4 accidents per 10,000 buses occur annually, with 1 in 44 traffic accidents leading to fatalities [59]. Thus, managing bus safety and assessing the driving risk of bus drivers is vital in transportation management. Driving risk scores will enable bus transit companies to evaluate the driving safety of their drivers and manage and train them to improve traffic safety. The recent advancements in sensing technology and legislation on the installation of in-vehicle sensors have enabled the real-time collection of driving records. Furthermore, the Traffic Safety Act Article 55 in South Korea, legislated in 2013, states that all commercial vehicles must be equipped with DTGs for traffic safety monitoring [60]. Besides, the ordinance (Article 30) of the Ministry of Land and Transport states that authorities should monitor ADBs through DTGs and use the records for managing vehicles, training drivers, and legislating traffic safety policies [61]. See Supplementary Material B for further details on ADBs and the use of DTGs in South Korea.

As described in Section II, existing studies on transportation safety have assessed the driving risk of vehicles, but they have considerable limitations. ADB frequencies have been measured from in-vehicle driving records, but these data were not aggregated into scores that may be easily understood by drivers [5], [62], [63]. Studies have not accurately assessed the driving risk of accident-free trips because of the small number of crash occurrence records. Therefore, in the present work, the authors have conducted driving risk assessment research with KOTSA to compute driving risk scores without using crash records. This section presents the results of applying the proposed method to actual driving records of buses in South Korea. Section IV-A describes the data and the implementation details of the proposed method. Sections IV-B and IV-C present the performance of the proposed method and the interpretation of the results, respectively.

A. Data and Driving Risk Assessment

The driving data used in the case study were obtained from a DTG database managed by KOTSA. The data consisted of DTG data of city buses operating in the southeastern region of the country from August to October 2019, during which the traffic volume of buses was high. During this period, 14,660 buses operated along 5,640 bus routes, which resulted in approximately 1,600,000 trips. We used the DTG data from 20 bus routes, each of which included approximately more than 2,000 trips in this period. The implementation details of the case study can be summarized as follows. In the ADB definition and preprocessing step, we used ten ADB definitions developed by KOTSA, which were defined based on the results of kinetics experiments and validated in road environments. Transportation safety researchers and managers in South Korea have used these ten ADB definitions as the

TABLE I
DESCRIPTIONS OF ADBs FOR BUSES DEFINED BY KOTSA

ADB	Bus criteria
Short-term over-speeding	Speed is 20 km/h above the speed limit
Long-term over-speeding	Speed is 20 km/h above the speed limit for 3 min
Rapid acceleration	Acceleration is greater than 6 km/h/s at a speed greater than 6 km/h
Rapid starting	Acceleration is greater than 8 km/h/s at a speed less than 5 km/h
Rapid deceleration	Deceleration is greater than 9 km/h/s at a speed greater than 6 km/h
Rapid stopping	Deceleration is greater than 9 km/h/s at a speed less than 5 km/h
Rapid lane changing	Change in rotation angle is 8°/s at a speed greater than 30 km/h, an acceleration (a deceleration) less than ±2 km/h/s, and an accumulated yaw rate change less than ±2°/s for 5 s
Rapid overtaking	Change in rotation angle is greater than 8°/s at a speed greater than 30 km/h, an acceleration greater than 3 km/h/s, and an accumulated yaw rate change less than ±2°/s for 5 s
Rapid turning	Accumulated change in rotation angle is from 60° to 120° (left or right) for 4 s at a speed greater than 20 km/h
Rapid U-turning	Accumulated change in rotation angle is from 160° to 180° (left or right) for 8 s at a speed greater than 15 km/h

standard reference [64], [65]. Refer to Supplementary Material B for further information. The corresponding details are listed in Table I.

Table II summarizes the ADB measurement data of the 20 bus routes. The ADB frequencies are affected by dynamic road elements, such as traffic flow and car-following situations along the route [66]. As buses follow designated routes, and trips on different routes have different ADB characteristics, we applied the proposed method to the ADB measurement data for each bus route to consider the influence of the driving route environment on the ADB pattern and driving risk score. The bus routes were coded as A to T to facilitate the analysis. Among the ten ADBs, the duration of long-term over-speeding was nearly zero, as city buses would stop at each bus station. Similarly, the duration of rapid starting was low because of the mechanical nature of city buses. Notably, the ADB duration differs according to the route. Route A showed a high average duration of rapid acceleration and short-term over-speeding at 20.03 and 31.08 (s), respectively. The average duration of rapid turns detected on Route F (60.01) was the highest among the 20 bus routes, whereas Route G had a particularly high average duration of short-term over-speeding (356.43).

During the preprocessing, the isolation forest (iForest) method was used to detect outliers in the ADB measurement data. iForest is an extension of the decision tree method based on the isolation mechanism, which is a procedure for separating outliers from the rest of the data through the iterative partitioning of the input space [67]. This method is faster than other methods [68]. We developed 100 iTrees containing 256 samples apiece. The proportion of outliers in the ADB measurement data on each route was set to 0.1, which was the basic setting used by [67]. In the latent risk factor extraction step, the ADB measurement data of 20 public transit buses were used as the input for the sparse NMF. The regularization parameter λ of the sparse NMF was set to 0.1. The non-negative double singular value decomposition method was used for initialization, as it can effectively initialize NMF algorithms with sparse factors and reduce the approximation error with rapid convergence [69].

B. Verification of the Driving Risk Score

Table III shows examples of ADB measurement and the driving risk score of the proposed method. The method can successfully compute high scores for the trips in which ADBs frequently occurred. Here, two trips with similar driving risk scores have different ADB distributions depending on the driving environment of the route. For example, upon comparing two trips with driving risk scores of approximately 70 at different routes, the trip at route D recorded 114 seconds of short-term over-speeding, whereas the trip at route E did not record short-term over-speeding. This result shows that the proposed method can successfully measure the risk level across different ADB distributions.

Meanwhile, no objectively accurate value is available for quantifying driving risk because of the unreliability and scarcity of accident-related records. As ADBs can be used to determine driving risk level, we developed an unsupervised-learning-based method to aggregate the ADB information of a trip into a driving risk score, and thus providing higher scores to trips in which ADBs frequently occur. Subsequently, the effectiveness of the proposed method can be verified by testing its ability to consistently distinguish trips with high-frequency ADBs and low-frequency ADBs through the risk scores. This experiment is analogous to the approach of [11] that assessed driving risk scores to determine if the scores could statistically differentiate safe and risky drivers.

In our experiment, the high- and low-risk groups were defined for each route as the sets of trips with driving risk scores in the top and bottom 25% respectively; previous studies used this categorization to validate their indicators or outcomes [70], [71]. If the proposed method can provide valid scores, then the average ADB durations will differ considerably between the two groups. We performed t-tests to examine the statistical difference between the average ADB durations of each group. If the p-value of the t-test is less than 0.05, the difference between the averages of the two groups is statistically significant. Refer to Supplementary Material D for the further results of verifying driving risk scores in the 5% and 10% risk groups.

Table IV lists the average ADB durations of the high- and low-risk groups on Routes A, B, and C identified using the proposed method. The results show the differences between two groups, with p-values of less than 0.05 for most ADBs. Thus, the differences between groups were significant for most ADBs. Large differences were obtained between the groups in terms of rapid acceleration and short-term over-speeding on Route A and rapid deceleration and rapid turning on Route B. The difference in terms of rapid acceleration was especially large on Route C. The driving risk score can be considered to reflect the different ADB characteristics on each route, as the

TABLE II

DETAILS OF THE ADB MEASUREMENT DATASET OF 20 SELECTED BUS SERVICE ROUTES

Route	Number of trips	Average ADB duration (s)									
		Rapid acceleration	Short-term over-speeding	Rapid deceleration	Rapid turning	Rapid stopping	Rapid lane changing	Rapid overtaking	Rapid U-turning	Rapid starting	Long-term over-speeding
A	3,020	20.03	31.08	6.69	12.34	1.13	0.5	0.01	2.44	0	0
B	3,117	4.78	2.04	4.92	9.89	1.76	2.4	0.01	2.99	0	0
C	1,987	22.21	0.11	9.99	11.02	1.48	1.4	0.24	2.41	0	0
D	11,904	41.26	85.27	17	11.2	1.73	0.76	0.11	14.27	0.8	0
E	17,489	34.32	0.19	10.76	8.15	2.74	0.52	0.02	5.85	0	0
F	4,073	15.63	2.84	8.82	60.01	2.17	4.3	0.19	4.68	0	0
G	3,581	148.67	356.43	87.2	20.76	9.57	0.83	0.65	4.07	0.1	0
H	4,386	14.82	0.49	11.1	11.92	2.75	1.37	0.02	3.65	0	0
I	6,709	15.13	9.01	14.16	9.83	3.45	0.99	0.08	4	0	0
J	7,308	37.76	0.04	3.83	12.11	0.74	0.51	0.02	14.15	0	0
K	4,834	16.45	0.47	4.4	9.58	2.05	1.64	0.01	2.07	0	0
L	4,141	4.18	12.71	6.69	21.58	1.46	2.89	0.01	1.13	0.17	0
M	2,124	8.03	7.94	5.61	11.75	0.94	0.9	0	2.23	0	0
N	2,230	9.74	4	8.1	9.12	1.44	0.94	0.01	0.76	0	0
O	2,642	7.44	0.56	6.73	6.16	1.26	1.75	0.01	0.175	0	0
P	3,880	6.73	2.77	4.13	17.72	0.63	8.99	0.01	2.29	0.4	0
Q	3,792	9.01	2.17	5.29	7.59	1.43	0.9	0	0.81	0.29	0
R	4,745	5.51	21.49	9.16	10.72	1.23	8.1	0.07	2.64	0	0
S	4,244	7.17	19.18	5.53	7.82	1.48	1.43	0.01	0.83	0	0
T	2,248	8.71	22.05	7.42	7.66	0.63	2.58	0.04	0.52	0	0

TABLE III

EXAMPLES OF ADB MEASUREMENT AND THE DRIVING RISK SCORE OF THE PROPOSED METHOD

Route	DR	Average ADB durations (s)									
		Rapid acceleration	Short-term over-speeding	Rapid deceleration	Rapid turning	Rapid stopping	Rapid lane changing	Rapid overtaking	Rapid U-turning	Rapid starting	Long-term over-speeding
D	**68.9**	7	114	11	14	1	0	0	0	0	0
	10.0	6	0	4	4	1	0	0	1	0	0
E	**72.9**	112	0	34	4	33	0	0	2	0	0
	10.0	19	0	2	1	0	0	0	0	0	0

TABLE IV

AVERAGE ADB DURATIONS FOR HIGH- AND LOW-RISK GROUPS (* DIFFERENCE BETWEEN THE HIGH- AND LOW-RISK GROUPS IS STATISTICALLY SIGNIFICANT AT THE 0.05 SIGNIFICANCE LEVEL)

Route	Groups	Rapid acceleration	Short-term over-speeding	Rapid deceleration	Rapid turning	Rapid stopping	Rapid lane changing	Rapid overtaking	Rapid U-turning	Rapid starting	Long-term over-speeding
A	**High-risk**	22.869	65.091	6.002	15.397	0.104	0.380	0.001	0.451	0	0
	Low-risk	4.333	0.009	5.015	5.661	1.067	0.492	0.001	1.272	0	0
	duration difference	18.536 *	65.082 *	0.987 *	9.736 *	-0.963 *	-0.112 *	0	-0.821 *	0	0
B	**High-risk**	7.556	1.156	8.061	18.423	1.244	2.803	0.003	3.963	0	0
	Low-risk	0.784	0.304	0.889	0.882	1.849	1.592	0	0.995	0	0
	duration difference	6.772 *	0.852 *	7.172 *	17.601 *	-0.605	1.211 *	0.003	2.968 *	0	0
C	**High-risk**	35.231	0.002	13.958	14.745	1.885	1.120	0.297	3.012	0	0
	Low-risk	6.462	0.005	2.700	5.948	0.225	1.271	0.09	1.228	0	0
	duration difference	28.769 *	-0.003	11.258 *	8.797 *	1.66 *	-0.151	0.207 *	1.784 *	0	0

differences between the high- and low-risk groups were large for the ADBs exhibiting high average durations on a given route (Table II). Thus, ADBs with high average durations and high variance are considered to be the primary ADBs affecting driving risk on a given route, and we regards them as distinct driving behaviors in determining driving risk.

The differences between some of the ADBs are marginal or even negative; that is, their average durations in the high-risk groups were less than those in the low-risk groups. However, in most of these cases, the t-test results indicate that the differences between these ADBs were not significant. As the ADBs occurred on trips in both groups, they were likely influenced by the environmental conditions along the route. Furthermore, the durations of these ADBs (e.g., rapid overtaking on Route A and short-term over-speeding on Route C) were low, and most trips on a given route did not exhibit such ADBs, further indicating that they did not affect the driving risk on these routes.

C. Interpretation of Latent Factors

As discussed in Section IV-B, the patterns of ADBs differ according to the bus route because the patterns are influenced by the traffic environment (e.g., traffic volume and road conditions). Therefore, the latent factors of the ADB dataset should result in distinct implications on each route. Here, the implications of the latent risk factors on the driving risk score were examined. The latent risk factors extracted via the

TABLE V
LOADINGS OF TEN ADBS ON LATENT RISK FACTORS

Route	Latent risk factors	Rapid acceleration	Short-term over-speeding	Rapid deceleration	Rapid turning	Rapid stopping	Rapid lane changing	Rapid overtaking	Rapid U-turning	Rapid starting	Long-term over-speeding
A	**1**	3,157.4	15,836.4	243.8	1,393.9	0.7372	0.8259	0.0001	5.8394	0	0
B	**1**	355.85	11.414	382.44	2,468.4	21.530	75.958	0	129.78	0	0
	2	242.01	7.5884	263.45	228.97	15.245	42.860	0.0368	86.534	0	0
C	**1**	5,697.7	0.0004	917.76	1,241.2	18.138	16.452	0.4746	57.896	0	0
	2	733.55	0	547.32	378.12	10.072	2.5321	8.0187	5.2734	0	0
	3	852.56	0.0003	331.79	943.05	49.444	11.642	7.9346	29.365	0	0

proposed method can reflect the distinct pattern of ADBs on each route.

Table IV lists the relative loadings of the ten ADBs on the latent risk factors on Routes A, B, and C, which were obtained from the basis matrix of the NMF. The numbers of latent risk factors were selected by considering the interpretability and validity of the NMF based on the MDL method [57], and they were one, two, and three for Routes A, B, and C, respectively. For the other routes, the maximum and minimum numbers of latent risk factors were four and one, respectively. Each route had different numbers of latent risk factors for explaining the ADB patterns. Therefore, the number of factors for determining driving risk differs by route, further implying that driving risk should be assessed separately for each route.

The loading of an ADB represents the influence of its frequency on the latent risk factor. Each latent risk factor can be explained in terms of the major ADBs that are noticeable on the route and in terms of other ADBs with marginal influence (Table V). The latent risk factor of Route A is characterized primarily by short-term over-speeding and rapid acceleration; that is, these are the ADBs that drivers on Route A engaged in most frequently on average. Other behaviors, such as rapid stopping, rapid lane changing, rapid overtaking, and rapid U-turns, resulted in small loadings on the latent risk factor and rarely occurred on Route A (Table II). On Route B, the first risk factor had the highest loading on rapid turning, whereas the second risk factor had high loadings on rapid acceleration, rapid deceleration, and rapid turning. Short-term over-speeding, rapid stopping, and rapid overtaking, which generally had small loadings on Route B, did not occur frequently. Similarly, the latent risk factors on Route C were influenced by the major ADBs of the route. The first risk factor was influenced by rapid turning; the second risk factor was influenced by rapid deceleration; and the third risk factor was influenced by rapid turning. Rapid acceleration resulted in high loadings on each latent risk factor on Route C.

Furthermore, the loadings of ADBs with a high variance were greater than those with a low variance, although the ADBs had similar durations. For example, rapid acceleration and short-term over-speeding occurred with similar durations on Route A, but the standard deviations of the duration of rapid acceleration and short-term over-speeding were 11.767 and 32.956, respectively. Thus, the loading of short-term over-speeding was higher than that of rapid acceleration. This implies that an ADB whose difference is large between drivers considerably affects the driving risk on a route.

V. COMPARISON WITH OTHER RISK ASSESSMENT METHODS

We evaluated the effectiveness of NMF for driving risk assessment by experimentally comparing its performance with PCA- and autoencoder-based assessments. We focused on evaluating the discrimination performance of high- and low-risk trips and the interpretability of the latent risk factors. As discussed in Section IV, the high- and low-risk groups were selected as the trips with driving risk scores in the top 25% and bottom 25%, respectively, for each method. We also performed t-tests to examine the statistical difference in the average ADB duration between risk groups.

For the next experiments, we adopted the settings used in the previous studies (Section II-B) that used the scoring methods based on PCA [38], [39]. For the PCA-based assessment, we first standardized the ADB dataset. We selected principal components that could explain more than 70% of the total variance and then calculated the driving risk score as the sum of the values of the selected components. Autoencoder-based scoring methods utilize various architectures suitable for the input characteristics and definitions of the scores. Here, the encoding matrix of the autoencoder was used to calculate the driving risk score because this matrix could indicate the values of driving risk factors. In view of learning the encoding matrix, the number of latent dimensions K in the autoencoder was selected to maximize the rate of decrease in the reconstruction error. In the comparative experiments, the layer architecture of the autoencoder was set to 10, $(10–K)/2$, K, $(10–K)/2$, and 10 neurons, with the hidden layers consisting of the average number of neurons in the input and code layers. The architecture of the decoder was a reverse of the encoder. The hyperbolic tangent function was used as the activation function between all layers. We used the mean squared error for the loss and the Adam optimizer to optimize the network parameters. The learning rate and number of training epochs were set to 0.0001 and 500, respectively.

The autoencoder uses nonlinear activation functions on each layer and does not output to the basis matrix; thus, the interpretability of its latent factors cannot be evaluated in terms of loadings. By contrast, in this work, we interpreted the outcomes of the PCA-based method. Table VI lists the loadings of the ten ADBs on the principal components extracted by the PCA method. Five principal components that could explain the variances of more than 70% were selected on Routes A, B, and C. Table VI shows that the PCA-based latent factors are difficult to interpret. As for latent factors of the proposed method, each ADB exhibited a non-negative influence on

TABLE VI

LOADINGS OF 10 ADBS ON LATENT RISK FACTORS OF PCA

Route	Latent risk factors	Rapid acceleration	Short-term over-speeding	Rapid deceleration	Rapid turning	Rapid stopping	Rapid lane changing	Rapid overtaking	Rapid U-turning	Rapid starting	Long-term over-speeding
A	**1**	-0.2065	-0.4261	0.3965	-0.2852	0.6053	0.0691	0.0359	0.4061	0	0
	2	0.4789	0.3655	0.4604	0.5196	0.1927	-0.1879	-0.0044	0.2874	0	0
	3	-0.0752	-0.0334	-0.1641	0.005	-0.0497	-0.4592	0.8514	0.1675	0	0
	4	-0.0312	0.1249	0.0095	0.2197	0.039	0.8496	0.4599	0.017	0	0
	5	0.5198	-0.1149	0.4137	-0.4409	-0.0383	0.0269	0.242	-0.5388	0	0
B	**1**	0.5179	0.1036	0.5145	0.4062	-0.2745	0.325	-0.0018	0.3322	0	0
	2	-0.275	-0.1176	-0.1393	-0.1686	-0.6454	0.6186	-0.0178	-0.2515	0	0
	3	-0.1351	0.9002	0.179	-0.0626	-0.0606	-0.0476	-0.2314	-0.2756	0	0
	4	-0.082	0.2207	0.0037	0.0848	0.0075	0.0352	0.9647	-0.0734	0	0
	5	-0.0466	0.2403	-0.2714	-0.3841	-0.3072	-0.1448	0.0426	0.7757	0	0
C	**1**	0.4605	-0.0222	0.5496	0.3349	0.5303	0.0063	0.1089	0.2831	0	0
	2	0.2978	-0.0432	-0.3965	0.4703	-0.4313	-0.0184	0.469	0.3533	0	0
	3	0.2177	0.2274	0.0329	-0.1907	0.0365	-0.7552	0.3885	-0.3754	0	0
	4	-0.0168	0.9245	-0.031	0.1777	-0.0566	-0.0051	-0.2859	0.166	0	0
	5	-0.0986	0.3005	0.1063	-0.3121	0.095	0.5534	0.6818	-0.1058	0	0

TABLE VII

AVERAGE ADB DURATIONS FOR HIGH- AND LOW-RISK GROUPS OBTAINED BY PCA (* THE DIFFERENCE BETWEEN THE HIGH- AND LOW-RISK GROUPS IS STATISTICALLY SIGNIFICANT AT THE 0.05 SIGNIFICANCE LEVEL)

Route	Groups	Rapid acceleration	Short-term over-speeding	Rapid deceleration	Rapid turning	Rapid stopping	Rapid lane changing	Rapid overtaking	Rapid U-turning	Rapid starting	Long-term over-speeding
A	**High-risk**	23.237	14.744	12.610	11.718	1.928	0.475	0.007	4.250	0	0
	Low-risk	10.302	22.718	1.086	10.932	0.100	0.165	0	0.268	0	0
	Duration difference	12.935 *	-7.974 *	11.524 *	0.786 *	1.828 *	0.310 *	0.007 *	3.982 *	0	0
B	**High-risk**	4.660	2.904	6.077	8.800	0.320	5.016	0.004	5.146	0	0
	Low-risk	3.078	0.061	2.359	8.291	4.854	0.394	0	1.232	0	0
	Duration difference	1.582 *	2.843 *	3.718 *	0.509	-4.534 *	4.622 *	0.004 *	3.914 *	0	0
C	**High-risk**	31.144	0.0268	12.513	13.902	1.746	0.989	0.641	3.119	0	0
	Low-risk	7.756	0	3.028	5.788	0.198	1.590	0	0.740	0	0
	Duration difference	23.388 *	0.0268 *	9.485 *	8.114 *	1.548 *	-0.601 *	0.641 *	2.379 *	0	0

TABLE VIII

AVERAGE ADB DURATIONS FOR HIGH- AND LOW-RISK GROUPS OBTAINED BY AUTOENCODER (* THE DIFFERENCE BETWEEN THE HIGH- AND LOW-RISK GROUPS IS STATISTICALLY SIGNIFICANT AT THE 0.05 SIGNIFICANCE LEVEL)

Route	Groups	Rapid acceleration	Short-term over-speeding	Rapid deceleration	Rapid turning	Rapid stopping	Rapid lane changing	Rapid overtaking	Rapid U-turning	Rapid starting	Long-term over-speeding
A	**High-risk**	10.116	4.305	2.867	7.969	0.846	0.528	0	0.089	0	0
	Low-risk	23.789	30.286	9.432	12.796	0.745	0.197	0.003	3.647	0	0
	Duration difference	-13.673 *	-25.981 *	-6.565 *	-4.827 *	0.101 *	0.331 *	-0.003	-3.558 *	0	0
B	**High-risk**	0.492	0.069	1.114	4.672	2.834	1.561	0	0.151	0	0
	Low-risk	7.314	2.132	6.605	10.693	0.800	2.270	0.004	6.336	0	0
	Duration difference	-6.822 *	-2.063 *	-5.491 *	-6.021 *	2.034 *	-0.709 *	-0.004	-6.185 *	0	0
C	**High-risk**	13.409	0	4.540	8.256	0.827	1.544	0.023	0.212	0	0
	Low-risk	26.781	0.019	8.686	12.332	0.726	0.833	0.496	3.975	0	0
	Duration difference	-13.372 *	-0.019	-4.146 *	-4.076 *	0.101 *	0.711 *	-0.473 *	-3.763 *	0	0

the driving risk score. In other words, ADBs with higher frequencies indicate higher latent factor values, which results in higher driving risk scores. However, the implications of the PCA-based latent factors for the driving risk score calculations are difficult to specify because the corresponding ADBs can have either positive or negative loadings on the principal components, and result in counterintuitive interpretation. For example, the driving risk score increases if the frequency of an ADB with a negative loading on a principal component decreases, which reveals that a decreasing ADB implies an increase in the driving risk. Therefore, the proposed method is more valid and useful than the PCA-based scoring method in terms of interpretability.

Tables VII and VIII list the average ADB durations of the high- and low-risk groups on Routes A, B, and C, as discriminated by the PCA and autoencoder methods, respectively. The results shown in Table VII indicate the poor performance of the PCA-based scoring method. The t-test results on risk groups also revealed that the difference in most of the ADBs was significant; however, the differences were smaller than those of the proposed method (Table IV). Specifically, the PCA-based driving risk score satisfactorily discriminated between trips with high and low ADB durations on Route C. However, the differences between the durations of ADBs in the high- and low-risk groups were not as large as those obtained using our method. For example, the gap between the high- and low-risk groups in terms of the PCA-based scores of rapid acceleration was smaller than that produced by the proposed method. Moreover, PCA could not distinguish the trips with high and low durations of ADBs on the other routes. On Route A, the average duration of short-term over-speeding in the low-risk group was greater than that in the high-risk group, while short-term over-speeding occurred the most frequently when driving on Route A. However, the trips with low PCA-based scores had higher frequencies of this ADB, which is an invalid result. The differences in the rapid acceleration durations were smaller than those obtained using our method, and the difference in the duration of rapid turning (i.e., the ADB that occurred the third most frequently) between the high- and low-risk groups was small. Similarly, PCA could not distinguish the major ADBs that frequently occurred on Route B. As each principal component was difficult to interpret semantically as a driving risk factor, the PCA-based scoring method, which simply defines driving risk as the sum of the values of the principal components, is not valid for ADB measurement.

The results in Table VIII indicate that the autoencoder-based scores cannot distinguish between high- and low-risk trips. According to the t-test results, the average frequencies of most ADBs in the low-risk group were considerably higher than those in the high-risk group in the autoencoder-based assessment. The autoencoder applies nonlinear transformation via activation functions in the neural network and determines latent representations to reconstruct the original data. However, this representation through complex nonlinear activation is unsuitable for calculating driving risk scores because it transforms the intrinsic relationship between ADB frequency and driving risk, and a higher ADB frequency corresponds to a higher driving risk score. Consequently, the autoencoder-based scores presented reversed relations to the ADB frequencies in the high- and low-risk groups. These results indicate that the proposed method is more valid than the PCA- and autoencoder-based assessment methods.

In summary, the proposed method outperformed other representation learning methods in terms of verifying driving risk and interpreting latent factors. The PCA-based scoring method led to information loss because only a few principal components were selected. Dense representations of latent factors with negative entries were extracted by the orthogonality assumption, and then they were interpreted counterintuitively in terms of ADBs and driving risk. Meanwhile, the autoencoder applied nonlinear activation functions to individual layers, consequently disrupting the positive relationships between the ADBs and driving risk scores. By contrast, the NMF reduced the information loss by minimizing the reconstruction error in the objective function. Non-negativity constraints enabled the NMF to satisfy the positive relationships between the ADBs and driving risk scores. The parts-based representation property originating from non-negativity enabled the combination of latent risk factors to explain the driving risk on a trip.

VI. Expert Evaluation

Two expert surveys were conducted during our research process, with the aim of improving our work and validating the contributions of our driving risk assessment method. In both surveys, we designed a questionnaire comprising evaluation criteria that could be scored on a seven-point Likert scale and open-ended questions that could evaluate the validity of the proposed work and subsequently improve the method. We recruited experts in academia (transportation researchers) and industry (e.g., transportation firm managers) with at least five years of expertise in their domains. The first survey was conducted in September 2020, in which the objective was to identify potential improvements in the research process. We performed a focus group interview with five experts and a survey with seven experts, who were requested to evaluate the validity and service applicability of the method in a safe-driving support system. The evaluation criteria of the first survey and their average scores are presented in Supplementary Material C. All experts evaluated positively the utility of the proposed method for providing safe-driving support services and discussed the considerations and potential scenarios for its application (Section VII). On this basis, we developed the regularized NMF for the clear interpretation of latent factors.

The second survey was conducted among 18 experts (academia and industry experts) in September 2021 to validate the key contributions of the proposed driving risk assessment method. The second survey was categorized into the following three sections to validate the motivation, result analysis, and contribution of our work: (1) representability of accident-related records versus ADBs on driving risk, (2) validity of the interpretation of latent risk factors, and (3) contribution of the proposed method.

In the first section, experts were requested to select a trip with an accident among a set of five trips in consideration of the duration of ten ADBs given by KOTSA and rank the five trips in the order of driving risk. We randomly selected a trip with an accident and four normal trips without an accident in each route. In other words, the durations of ten ADBs of the five trips in the four routes were presented to each subject. Refer to Supplementary Material C for an example of the questionnaire items. The average accuracy of identifying the trip with an accident was 0.014, which is equivalent to 1/72. Thus, the experts could not identify the trip with an accident. Moreover, we evaluated the correlation and distance between the rankings determined by the experts and those by the proposed method in the order of driving risk via Spearman rank correlation and Levenshtein distance [72], [73]. The resultant average Levenshtein distance was 3.222, which indicates that two risk rankings can be equivalent with roughly two operations. The average rank correlation was 0.356 for all risk rankings evaluated by the experts, which denotes that the driving risk score is consistent with the risk perception of experts on the ADB duration. This result supports the findings of previous studies that ADBs are major indicators for assessing driving risk in a trip, whereas accident records are less useful as a measure of driving risk.

In the second section, experts were requested to evaluate whether the interpretation of latent factors based on the loading matrix of the ADB measurement data was valid and subsequently interpret the factors. We offered the loading matrices

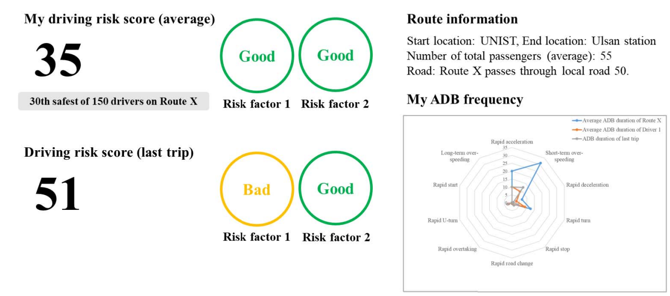


Fig. 3. Use case of risk assessment of a driver.

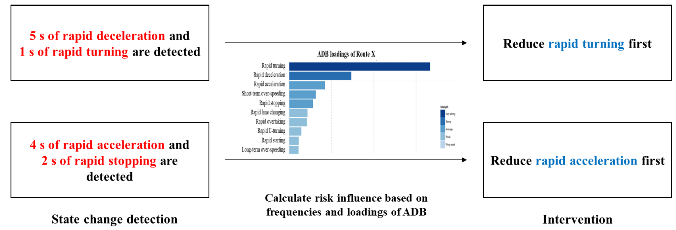


Fig. 4. Example of the intervention in practice.

of the ADB measurement data, average durations of ADBs, and our interpretation based on the loading matrix for the three routes presented in Section IV. Refer to Supplementary Material C for our interpretation of latent factors. The experts evaluated the validity of our interpretation of latent factors as 5.125, 5.625, and 5.375 on average. Moreover, the interpretation of latent risk factors by experts corresponded to our interpretation. As such, the latent factors extracted from the proposed method are valid, and they are useful for interpreting and managing driving risk.

During their deliberation, first, the experts reached an agreement that accident-related records other than those involving serious accidents are difficult to acquire. They also found that latent risk factors that consider the interaction among ADBs are meaningful and applicable in practice. Most experts viewed the method of developing a driving risk assessment method for assessing the driving risks of non-accident trips as critical. Second, we found that some experts in practice still emphasized the utility of accident-related records but admitted the scarcity and the bias of those records. An expert claimed that ADBs and accident-related records are complementary measures for determining driving risk. Although the first finding concurs with the literature and our work that highlight the necessity of risk assessment not predicated by accident-related records [13], the second finding indicates that accident-related records are still useful in practice.

VII. Implementation of the Proposed Method

In both surveys and interviews, the experts discussed three common considerations that should be addressed when implementing the proposed method in the future: (1) applicability of the method, (2) interventions for driving risk management based on the method, and (3) necessity of using reliable in-vehicle sensors for accurate ADB measurements.

According to all experts, the development of driving risk assessment methods that do not require the use of accident-related records has been a significant challenge in the transportation field. An expert emphasized that the proposed method can assess the risk level of a driver because it can assess a trip as the basic unit of analysis, i.e., by averaging the driving risk scores of all trips that the driver operated. Fig. 3 shows an example of the use case of assessing the risk level of a driver. As suggested by the expert, the proposed method will be able to obtain effective risk information of drivers if combined with the geographic properties of the route, but this type of data is not recorded in the DTG database. By using methods that do not require such records, transportation companies can evaluate the driving safety of even employees who have not been in crashes and subsequently train these employees to reduce their risky driving behaviors, which contributes to overall traffic safety. Another expert shared that local governments may utilize the driving risk score when providing subsidies (e.g., for fuel, operational loss, and administration) to individual drivers, companies, and local governments. Similarly, insurance companies may provide premium discounts based on driver risk scores via pay-as-you-drive schemes to encourage both safe and risky drivers to drive safely.

The proposed method can provide guidelines to reduce driving risk based on the frequencies and loadings of ADBs. Each loading w_{ij} of the basis matrix W represents the influence of ADB i on the risk factor j. As the driving risk score of a trip is defined as the average of standardized latent factor values of the trip, the risk influence of an ADB on the risk score is proportional to the average of all loadings of the ADB to the latent factors on the route. Meanwhile, the total risk influence of an ADB can be defined as the product of the duration and the risk influence of the ADB. The proposed method can provide various levels of notifications to drivers as they perform specific ADBs based on the total risk influence (Fig. 4). This notification can be delivered in real time during driving in accordance with the state changes of drivers. Thus, the proposed learning-based approach can identify clear levels of ADB frequencies that can be effectively used to monitor, evaluate, and control the state changes of drivers in real time. An expert also revealed that interventions integrating the quantitative influences with additional trip information may effectively encourage drivers to reduce aggressive driving. The spatial and temporal information of a driver who has driven aggressively on the route can be used to provide guidance to drivers, thus preventing ADBs in advance.

Accurate ADB measurement is indispensable for ADB-based driving risk assessment. The ADB dataset used in our model was based on measurements obtained from in-vehicle sensors. Some experts mentioned that in-vehicle sensor records include noise and measurement errors. In an experiment that compared the differences in speed, RPM, azimuth, and break measurements among ten types of DTGs, the differences were attributed to errors from various sensors and manufacturing processes [74]. The GPS is typically unable to provide accurate and reliable locations because of the presence of signal interruptions or blockage during driving [75], [76]. As a means of overcoming signal blockage, the inertial navigation system (INS) or dead reckoning (DR) are commonly integrated into GPS units. However, both INSs and DR can cause measurement biases in the inertial sensors [77]. Erroneous ADB measurements from noisy sensors contain distorted patterns of ADB frequencies that may produce inaccurate driving risk scores. The accuracy and reliability of in-vehicle sensors should be ensured to prevent erroneous ADB measurements.

VIII. Conclusion

Aggressive driving is a major cause of disturbance to traffic safety. ADB-based driving risk assessment is an effective tool for transportation companies and governments for managing driving safety. In this study, we proposed an NMF-based driving risk assessment method that can aggregate the frequencies of ADBs over the course of a trip. In this method, interpretable latent risk factors from ADB measurements are extracted and driving risk scores based on the aggregation of latent risk factor values are calculated. We applied the method to actual driving records of buses on multiple routes. A real-world application and comparative experiments with other risk assessment methods were demonstrated, and the superiority of the proposed method in terms of discriminating high- and low-risk trips and interpreting driving risk factors was evaluated. The acceptability and applicability of the proposed method was validated by surveys with experts from academia and industry.

The proposed method is the first one to assess driving risk by using only driving behavior records. We used ADBs as surrogate measures of driving risk, and trips with high or low frequencies of ADB indicate high- or low-risk driving. The proposed method can scientifically assess the driving risks of regular trips in which no crashes occur, and it does not rely on accident-related records, which are typically unreliable and difficult to acquire. Sparse NMF is superior to the existing risk assessment methods in terms of both the discrimination and interpretation of driving risk, rendering it suitable for use in service applications for providing safe-driving support. The proposed method can be used to help drivers monitor their driving behaviors and reduce ADBs by providing feedback for lowering their driving risk scores. Such driving evaluation and management are necessary for various stakeholders in the transportation domain.

Meanwhile, the study has some limitations that can be addressed in the future. One of the limitations is that the proposed method could not discriminate some minor ADBs on a route. Marginal errors are sometimes unavoidable when using an unsupervised learning approach for risk assessment. Our method was able to aggregate the frequencies of ADBs into a risk score by highlighting the weighted effects of major ADBs, but this approach resulted in unsuccessful discrimination of some minor ADBs. The other limitation is that the ten ADBs utilized in the experiments were the fixed threshold-based. Although those ADBs were defined and validated through experiments in actual road environments by KOTSA, the threshold-based ADB criteria have not considered the situations in which ADBs adaptively affect driving risk depending on the environment. In future research, we may develop driving risk assessment methods that can adaptively consider ADBs with spatial and temporal trip information and the various driving conditions. Finally, on the basis of advice from experts from the field who highlighted the importance of accident-related records, a self-supervised learning approach may be adopted to extract the latent risk factors of ADB measurements, thereby obtaining a high-risk score for a trip with high-frequency ADBs and accident-related labels.

Acknowledgment

The authors would like to thank the associate editor and anonymous reviewers for their helpful and insightful comments that have resulted in a much improved version of this manuscript.

References

[1] National Highway Traffic Safety Administration, U.S. Department of Transportation, and Traffic Tech. (Jul. 2015). *Traffic Safety Facts 2013 Data*. [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812169

[2] L. Evans, *Traffic Safety*, vol. 179. Lansing, MI, USA: Science Serving Society, 2004.

[3] R. Tian, L. Li, M. Chen, Y. Chen, and G. J. Witt, "Studying the effects of driver distraction and traffic density on the probability of crash and near-crash events in naturalistic driving environment," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1547–1555, Sep. 2013.

[4] National Highway Traffic Safety Administration, U.S. Department of Transportation, and Traffic Tech. (1998). *Aggressive Drivers View Traffic Differently, Capital Bellway Focus Groups Find*. [Online]. Available: https://one.nhtsa.gov/people/outreach/traftech/pub/tt175.html

[5] J. Lee and J. Kitae, "A framework for evaluating aggressive driving behaviors based on in-vehicle driving records," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 65, pp. 610–619, Aug. 2019, doi: 10.1016/j.trf.2017.11.021.

[6] H. Zhang, W. Qu, Y. Ge, X. Sun, and K. Zhang, "Effect of personality traits, age and sex on aggressive driving: Psychometric adaptation of the driver aggression indicators scale in China," *Accident Anal. Prevention*, vol. 103, pp. 29–36, Jun. 2017.

[7] T. Toledo, O. Musicant, and T. Lotan, "In-vehicle data recorders for monitoring and feedback on drivers' behavior," *Transp. Res. C, Emerg. Technol.*, vol. 16, no. 3, pp. 320–331, 2008.

[8] M. Islam and F. Mannering, "A temporal analysis of driver-injury severities in crashes involving aggressive and non-aggressive driving," *Anal. Methods Accident Res.*, vol. 27, Sep. 2020, Art. no. 100128.

[9] J. Paefgen, T. Staake, and E. Fleisch, "Multivariate exposure modeling of accident risk: Insights from pay-as-you-drive insurance data," *Transp. Res. A, Policy Pract.*, vol. 61, pp. 27–40, Mar. 2014.

[10] J. Wang, Y. Zheng, X. Li, C. Yu, K. Kodaka, and K. Li, "Driving risk assessment using near-crash database through data mining of tree-based model," *Accident Anal. Prevention*, vol. 84, pp. 54–64, Nov. 2015, doi: 10.1016/j.aap.2015.07.007.

[11] T. Osafune, T. Takahashi, N. Kiyama, T. Sobue, H. Yamaguchi, and T. Higashino, "Analysis of accident risks from driving behaviors," *Int. J. Intell. Transp. Syst. Res.*, vol. 15, no. 3, pp. 192–202, 2017.

[12] F. L. Mannering and C. R. Bhat, "Analytic methods in accident research: Methodological frontier and future directions," *Anal. Methods Accident Res.*, vol. 1, pp. 1–22, Jan. 2014.

[13] L. Zheng, T. Sayed, and F. Mannering, "Modeling traffic conflicts for use in road safety analysis: A review of analytic methods and future directions," *Anal. Methods Accident Res.*, vol. 29, Mar. 2021, Art. no. 100142.

[14] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[15] L. Tasca, "A review of the literature on aggressive driving research," Ontario Advisory Group Safe Driving Secretariat, Road User Saf. Branch, Ontario Ministry Transp., 2000. [Online]. Available: https://books.google.co.kr/books?id=ZWnRoAEACAAJ

[16] R. Ivers *et al.*, "Novice drivers' risky driving behavior, risk perception, and crash risk: Findings from the DRIVE study," *Amer. J. Public Health*, vol. 99, no. 9, pp. 1638–1644, Sep. 2009.

[17] A. Lambert-Bélanger, S. Dubois, B. Weaver, N. Mullen, and M. Bédard, "Aggressive driving behaviour in young drivers (aged 16 through 25) involved in fatal crashes," *J. Saf. Res.*, vol. 43, nos. 5–6, pp. 333–338, Dec. 2012, doi: 10.1016/j.jsr.2012.10.011.

[18] D. Shinar and R. Compton, "Aggressive driving: An observational study of driver, vehicle, and situational variables," *Accident Anal. Prevention*, vol. 36, no. 3, pp. 429–437, May 2004.
[19] M. Abou-Zeid, I. Kaysi, and H. Al-Naghi, "Measuring aggressive driving behavior using a driving simulator: An exploratory study," in *Proc. 3rd Int. Conf. Road Saf. Simulation*, 2011, pp. 1–19.
[20] M. Čabarkapa, M. Čubranić-Dobrodolac, S. Čičević, and B. Antić, "The influence of aggressive driving behavior and impulsiveness on traffic accidents," *Int. J. Traffic Transp. Eng.*, vol. 8, no. 3, pp. 1–18, 2018.
[21] A. B. R. Gonzalez, M. R. Wilby, J. J. V. Diaz, and C. S. Ávila, "Modeling and detecting aggressiveness from driving signals," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 4, pp. 1419–1428, Aug. 2014.
[22] A. Jahangiri and H. A. Rakha, "Applying machine learning techniques to transportation mode recognition using mobile phone sensor data," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2406–2417, Oct. 2015.
[23] Y. Li, C. Miyajima, N. Kitaoka, and K. Takeda, "Measuring aggressive driving behavior using signals from drive recorders," in *Proc. 17th IEEE Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2014, pp. 1886–1887.
[24] Y. Park, "The trend of dangerous driving behaviour criteria," *Auto J., J. Korean Soc. Automot. Eng.*, vol. 39, no. 9, pp. 38–42, 2017.
[25] F. Guo and Y. Fang, "Individual driver risk assessment using naturalistic driving data," *Accident Anal. Prevention*, vol. 61, no. 6, pp. 3–9, Dec. 2013, doi: 10.1016/j.aap.2012.06.014.
[26] J. Chen, Z. Wu, and J. Zhang, "Driving safety risk prediction using cost-sensitive with nonnegativity-constrained autoencoders based on imbalanced naturalistic driving data," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 12, pp. 4450–4465, Dec. 2019.
[27] T. E. Boyce and E. S. Geller, "An instrumented vehicle assessment of problem behavior and driving style: Do younger males really take more risks?" *Accident Anal. Prevention*, vol. 34, no. 1, pp. 51–64, Jan. 2002.
[28] F. Guo, S. G. Klauer, J. M. Hankey, and T. A. Dingus, "Near crashes as crash surrogate for naturalistic driving studies," *Transp. Res. Rec.*, vol. 2147, no. 1, pp. 66–74, 2010.
[29] T. A. Dingus *et al.*, "The 100-car naturalistic driving study, phase II-results of the 100-car field experiment," NHSTA, Washington, DC, USA, Tech. Rep. DOT HS 810 593, Apr. 2006.
[30] J. K. Jonasson and H. Rootzén, "Internal validation of near-crashes in naturalistic driving studies: A continuous and multivariate approach," *Accident Anal. Prevention*, vol. 62, pp. 102–109, Jan. 2014.
[31] E. G. Mantouka, E. N. Barmpounakis, and E. I. Vlahogianni, "Identifying driving safety profiles from smartphone data using unsupervised learning," *Saf. Sci.*, vol. 119, pp. 84–90, Nov. 2019.
[32] G. Li, Y. Chen, D. Cao, X. Qu, B. Cheng, and K. Li, "Extraction of descriptive driving patterns from driving data using unsupervised algorithms," *Mech. Syst. Signal Process.*, vol. 156, Jul. 2021, Art. no. 107589.
[33] M. M. Hossain, X. Sun, E. Mitran, and M. A. Rahman, "Investigating fatal and injury crash patterns of teen drivers with unsupervised learning algorithms," *IATSS Res.*, vol. 45, no. 4, pp. 561–573, Dec. 2021.
[34] M.-J. Kim, C.-H. Lim, C.-H. Lee, K.-J. Kim, Y. Park, and S. Choi, "Approach to service design based on customer behavior data: A case study on eco-driving service design using bus drivers' behavior data," *Service Bus.*, vol. 12, no. 1, pp. 203–227, 2018.
[35] T. Li, H. Zhang, C. Yuan, Z. Liu, and C. Fan, "A PCA-based method for construction of composite sustainability indicators," *Int. J. Life Cycle Assessment*, vol. 17, no. 5, pp. 593–603, 2012.
[36] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 2, no. 4, pp. 433–459, 2010.
[37] M. Tschannen, O. Bachem, and M. Lucic, "Recent advances in autoencoder-based representation learning," 2018, *arXiv:1812.05069*.
[38] L. Anderloni, E. Bacchiocchi, and D. Vandone, "Household financial vulnerability: An empirical analysis," *Res. Econ.*, vol. 66, no. 3, pp. 284–296, Sep. 2012, doi: 10.1016/j.rie.2012.03.001.
[39] J. Choi, Y. S. Park, and J. D. Park, "Development of an aggregate air quality index using a PCA-based method: A case study of the U.S. transportation sector," *Amer. J. Ind. Bus. Manage.*, vol. 5, no. 2, pp. 53–65, 2015.
[40] L. Jia, W. Zhang, R. Jia, H. Zhang, and X. Chen, "Construction formula of biological age using the principal component analysis," *BioMed Res. Int.*, vol. 2016, Dec. 2016, Art. no. 4697017.
[41] W. G. Zhang *et al.*, "Construction of an integral formula of biological age for a healthy Chinese population using principle component analysis," *J. Nutrition, Health Aging*, vol. 18, no. 2, pp. 137–142, 2014.
[42] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.
[43] D. Vilenchik, B. Yichye, and M. Abutbul, "To interpret or not to interpret PCA? This is our question," in *Proc. Int. AAAI Conf. Web Social Media*, vol. 13, 2019, pp. 655–658.
[44] T.-N. Nguyen, H.-H. Huynh, and J. Meunier, "Estimating skeleton-based gait abnormality index by sparse deep auto-encoder," in *Proc. IEEE 7th Int. Conf. Commun. Electron. (ICCE)*, Jul. 2018, pp. 311–315.
[45] T.-N. Nguyen and J. Meunier, "Applying adversarial auto-encoder for estimating human walking gait abnormality index," *Pattern Anal. Appl.*, vol. 22, no. 4, pp. 1597–1608, Nov. 2019, doi: 10.1007/s10044-019-00790-7.
[46] F. Xu, Z. Huang, F. Yang, D. Wang, and K. L. Tsui, "Constructing a health indicator for roller bearings by using a stacked auto-encoder with an exponential function to eliminate concussion," *Appl. Soft Comput.*, vol. 89, Apr. 2020, Art. no. 106119, doi: 10.1016/j.asoc.2020.106119.
[47] C. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 45–55, Jan. 2010.
[48] L. Lee and D. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2001, pp. 556–562.
[49] L. Eldén, *Matrix Methods in Data Mining and Pattern Recognition*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2007.
[50] Y. Wang, Y. Jia, C. Hu, and M. Turk, "Non-negative matrix factorization framework for face recognition," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 19, no. 4, pp. 495–511, Jun. 2005.
[51] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in *Proc. 26th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2003, pp. 267–273.
[52] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111–126, 1994.
[53] K.-H. Kim, K.-J. Kim, C. Lim, and J.-Y. Heo, "Development of a lifelogs-based daily wellness score to advance a smart wellness service," *Service Sci.*, vol. 10, no. 4, pp. 408–422, 2018.
[54] D. A. Sass, M. P. Twohig, and W. H. Davies, "Defining the independent variables and ensuring treatment integrity: A comparison across journals of different theoretical orientations," *Behav. Therapist*, vol. 27, no. 8, pp. 172–174, Winter 2004.
[55] *Prevalence of Self-Reported Aggressive Driving Behavior: United States, 2014*, AAA Found. Traffic Saf., Washington, DC, USA, 2016, p. 11.
[56] P. O. Hoyer, "Non-negative sparse coding," in *Proc. 12th IEEE Workshop Neural Netw. Signal Process.*, Jan. 2002, pp. 557–565.
[57] S. Squires, A. Prügel-Bennett, and M. Niranjan, "Rank selection in nonnegative matrix factorization using minimum description length," *Neural Comput.*, vol. 29, no. 8, pp. 2164–2176, Aug. 2017.
[58] C. Ma, W. Hao, W. Xiang, and W. Yan, "The impact of aggressive driving behavior on driver-injury severity at highway-rail grade crossings accidents," *J. Adv. Transp.*, vol. 2018, Oct. 2018, Art. no. 9841498.
[59] B. Choi, J. Kang, S. Kim, and S. Kwak, "One of every eleven buses crashes every year," Korea Transp. Saf. Authority, Republic of Korea, Tech. Rep. 1, Mar. 2018.
[60] (2013). *Traffic Safety Act*. [Online]. Available: https://www.law.go.kr
[61] (2018). *Ordinance of the Ministry of Land and Transport Article 30*. [Online]. Available: https://www.law.go.kr/
[62] M.-J. Kim, C. Lim, and K.-J. Kim, "A data-driven approach to designing new services for vehicle operations management," *Int. J. Ind. Eng.*, vol. 25, no. 5, pp. 604–619, 2018.
[63] J.-M. Jang, Y.-I. Lee, and J.-K. Lim, "Analysis of risk driver discrimination considering driver's behavior based on digital TachoGraph (DTG)," *J. Transp. Res.*, vol. 24, no. 4, pp. 63–77, 2017.
[64] E. M. Park and J. Kang, "Analysis of dangerous driving behavior by bus type using digital tacho graph in Sejong City," *J. Korean Soc. Transp.*, vol. 38, no. 6, pp. 462–472, 2020.
[65] J. Cho, K. Jang, J. Shim, and C. Oh, "An analysis of dangerous driving behaviors in commercial vehicles using eTAS," in *Proc. KOR-KST Conf.*, vol. 72. Sejong, South Korea: Korean Society of Transportation, 2015, pp. 158–163.
[66] L. Yan, Z. Huang, Y. Zhang, L. Zhang, D. Zhu, and B. Ran, "Driving risk status prediction using Bayesian networks and logistic regression," *IET Intell. Transport Syst.*, vol. 11, no. 7, pp. 431–439, Sep. 2017.
[67] F. T. Liu, K. M. Ting, and Z. H. Zhou, "Isolation forest," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Dec. 2008, pp. 413–422.
[68] Y. Chen and W. Wu, "Isolation forest as an alternative data-driven mineral prospectivity mapping method with a higher data-processing efficiency," *Natural Resour. Res.*, vol. 28, no. 1, pp. 31–46, 2019, doi: 10.1007/s11053-018-9375-6.

[69] C. Boutsidis and E. Gallopoulos, "SVD based initialization: A head start for nonnegative matrix factorization," *Pattern Recognit.*, vol. 41, no. 4, pp. 1350–1362, Apr. 2008.

[70] S. S. Hayek *et al.*, "Soluble urokinase receptor and chronic kidney disease," *New England J. Med.*, vol. 373, no. 20, pp. 1916–1925, 2015.

[71] S. L. Stacy *et al.*, "Perinatal outcomes and unconventional natural gas operations in southwest Pennsylvania," *PLoS ONE*, vol. 10, no. 6, 2015, Art. no. e0126425.

[72] J. H. Zar, "Spearman rank correlation," in *Encyclopedia of Biostatistics*, vol. 7. Wiley, 2005. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/0470011815.b2a15150

[73] L. Yujian and L. Bo, "A normalized Levenshtein distance metric," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1091–1095, Jun. 2007.

[74] J. Park, G. Joh, and J. Park, "Study on reliability of new digital tachograph for traffic accident investigation and reconstruction," *Trans. Korean Soc. Automot. Eng.*, vol. 23, no. 6, pp. 615–622, Nov. 2015.

[75] K. Jo, K. Chu, and M. Sunwoo, "Interacting multiple model filter-based sensor fusion of GPS with in-vehicle sensors for real-time vehicle positioning," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 329–343, Mar. 2012.

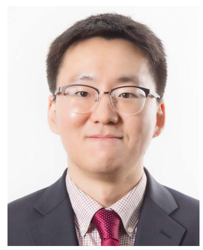
[76] H. Sim, C. Chae, M. Kang, and J. Lee, "Analysis on accuracy of GPS installed in digital tachograph of commercial vehicles," *J. Korea Inst. Intell. Transp. Syst.*, vol. 18, no. 6, pp. 164–175, Dec. 2019.

[77] X. Song, X. Li, W. Tang, and W. Zhang, "A fusion strategy for reliable vehicle positioning utilizing RFID and in-vehicle sensors," *Inf. Fusion*, vol. 31, pp. 76–86, Sep. 2016.

Hyunwoo Seo (Student Member, IEEE) received the B.S. degree from the Department of Mathematical Science, Ulsan National Institute of Science and Technology (UNIST), Republic of Korea, in 2020, where he is currently pursuing the Ph.D. degree with the Department of Industrial Engineering. His research interest includes knowledge discovery through representation learning and self-supervised learning.

Jongkyung Shin (Student Member, IEEE) received the B.S. degree from the Department of Industrial Engineering, UNIST, in 2020, where he is currently pursuing the Ph.D. degree with the Graduate School of Artificial Intelligence. His research interests include machine learning and AI-based service systems.

Ki-Hun Kim received the B.S. and Ph.D. degrees from the Department of Industrial and Management Engineering, Pohang University of Science and Technology (POSTECH), Republic of Korea. He is currently an Assistant Professor with the Department of Industrial Engineering, Pusan National University, Republic of Korea. His research interest includes industrial data analytics to address industrial problems. His work has been applied in various areas, such as healthcare, transportation, and manufacturing.



Chiehyeon Lim (Member, IEEE) received the B.S. and Ph.D. degrees from the Department of Industrial and Management Engineering, POSTECH. As part of his postdoctoral research experience, he worked as an Assistant Project Scientist and a Lecturer with the School of Engineering, UC Merced. He is currently an Associate Professor with the Department of Industrial Engineering and the Graduate School of Artificial Intelligence, UNIST. His research interests include service engineering and knowledge discovery methods.

Jungcheol Bae received the B.S. degree from the Department of Psychology, Yeongnam University. He is currently the Seoul Area Chief Director of KOTSA. His research interest includes driver safety behavior analysis by using the big data and intelligent traffic safety information service.