Travel Demand Forecasting: A Fair AI Approach

Xiaojian Zhang, Qian Ke, and Xilei Zhao

***Abstract*— Artificial Intelligence (AI) and machine learning have been increasingly adopted for travel demand forecasting. The AI-based travel demand forecasting models, though generate accurate predictions, may produce prediction biases and raise fairness issues. Using such biased models for decision-making may lead to transportation policies that exacerbate social inequalities. However, limited studies have been focused on addressing the fairness issues of these models. Therefore, in this study, we propose a novel methodology to develop fairness-aware, highly-accurate travel demand forecasting models. Particularly, the proposed methodology can enhance the fairness of AI models for multiple protected attributes (such as race and income) simultaneously. Specifically, we introduce a new fairness regularization term, which is explicitly designed to measure the correlation between prediction accuracy and multiple protected attributes, into the loss function of the travel demand forecasting model. We conduct two case studies to evaluate the performance of the proposed methodology using real-world ridesourcing-trip data in Chicago, IL and Austin, TX, respectively. Results highlight that our proposed methodology can effectively enhance fairness for multiple protected attributes while preserving prediction accuracy. Additionally, we have compared our methodology with three state-of-the-art methods that adopt the regularization term approach, and the results demonstrate that our approach significantly outperforms them in both preserving prediction accuracy and enhancing fairness. This study can provide transportation professionals with a new tool to achieve fair and accurate travel demand forecasting.**

***Index Terms*— AI, fairness, forecasting, machine learning, regularization, travel demand.**

I. Introduction

IN RECENT years, Artificial Intelligence (AI) has been increasingly used in travel behavior analysis, due to its powerful prediction capability [1], [2], [3], [4]. However, a growing number of studies reported that AI has evident fairness issues [5], [6], [7], [8], [9], [10]—making worse predictions for disadvantaged population groups (e.g., racial and ethnic underrepresented groups, low-income individuals, and women) than the advantaged groups. The unfairness of AI may negatively impact transportation policies and decision-making. For example, research has indicated that AI predictive biases can skew the allocation of safety improvement grants towards advantaged communities and cause the road infrastructure in disadvantaged communities to receive fewer investments [11]. Such unfairness issues are found to be more pronounced in AI-based travel demand forecasting models [12], [13]. Specifically, unfair predictions of travel demand may cause ridesourcing operators to consider reallocating fewer vehicles [14] to the disadvantaged neighborhoods, consequently leading to higher surge pricing [15], higher per-mile fees [16] and longer waiting times [17] in these areas. Additionally, unfair travel demand predictions may misinform transit agencies and city governments in determining investments for adding new transit services or infrastructures in disadvantaged communities [13]. The unfair travel demand predictions may also cause AI-based traffic management systems to prioritize traffic flows in ways that benefit advantaged areas over disadvantaged ones, which may further worsen traffic congestion in disadvantaged communities [18]. Over time, these ill-informed transportation policies or decision-making caused by unfair travel demand predictions could lead to further unintended consequences for transportation equity [13], [19]. Accordingly, it is crucial to incorporate fairness into AI-based travel demand forecasting models [20].

Recently, some researchers have started to develop fairness-aware AI methods in travel behavior modeling [11], [12], [13], [21]. However, research on this important topic, especially for travel demand forecasting (in this paper, we refer to the first step of the four-step travel demand model, namely, trip generation), is still lacking. For instance, although various methods have been developed to mitigate the unfairness issues, very few can be flexibly adopted by different types of models (e.g., linear models, deep learning models with different architectures, etc.). In other words, there still lacks a systematic framework to address the model's fairness issue in a model-agnostic (i.e., the method should be independent of models) manner. Also, it remains largely unsolved how to prioritize model fairness while preserving its prediction accuracy, both of which are critical to ensure the trustworthiness of AI [22], [23]. Additionally, previous studies have primarily focused on correcting the unfairness of a single protected attribute. In real-world dataset, however, the debiased model and results could vary across different protected attributes, potentially causing confusion and hindering adoption by end-users. For example, one study has found that mitigating unfairness of one protected attribute (i.e., race) could increase the prediction disparities of another protected attribute (i.e., income) [13]. This suggests that a model that is fair for one protected attribute could still be unfair for other attributes [24]. However, few prior studies

Manuscript received 23 September 2023; revised 11 March 2024; accepted 15 April 2024. This work was supported by the U.S. Department of Transportation through the Southeastern Transportation Research, Innovation, Development and Education (STRIDE), Region 4 University Transportation Center under Grant 69A3551747104 and through the Tier 1 University Transportation Center, the Center for Equitable Transit-Oriented Communities (CETOC) under Grant 69A3552348337. The authors also thank Bloomberg for supporting this research. The Associate Editor for this article was P. Wang. *(Corresponding author: Xiaojian Zhang.)*

Xiaojian Zhang and Xilei Zhao are with the Department of Civil and Coastal Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: xiaojianzhang@ufl.edu; xilei.zhao@essie.ufl.edu).

Qian Ke is with Bloomberg, New York, NY 10022 USA (e-mail: qke4@bloomberg.net).

This article has supplementary downloadable material available at https://doi.org/10.1109/TITS.2024.3395061, provided by the authors.

Digital Object Identifier 10.1109/TITS.2024.3395061

have been devoted to simultaneously tackling fairness issues from multiple protected attributes [24], [25].

To address these research gaps, we aim to develop a new methodology to enhance fairness in AI-based travel demand forecasting models, especially focusing on trip generation forecasting. More specifically, first, we define *Fairness* as the *Equality of Prediction Accuracy*, i.e., the prediction accuracy is equal for advantaged and disadvantaged population groups. Next, we examine the potential unfairness (i.e., prediction accuracy disparity) existing among several state-of-the-art deep learning and statistical models for travel demand forecasting, using real-world ridesourcing-trip data in Chicago, IL and Austin, TX. We propose a novel absolute correlation regularization method to simultaneously correct the detected unfairness across multiple protected attributes (e.g., race, education, etc). We further compare the proposed methodology with other existing state-of-the-art regularization terms to show its effectiveness in both preserving accuracy and correcting unfairness. The unique contributions of this study are presented as follows:

- This study is one of the first studies to examine the fairness issues of travel demand forecasting models from the algorithmic view. We improve knowledge on this topic by detecting the unfairness issues of several commonly-used deep learning and statistical models and proposing a methodology to correct the unfairness.
- We introduce a novel absolute correlation regularization term to address the model's unfairness arising from *multiple protected attributes*. This regularization term is explicitly designed to penalize models that produce unfair predictions, which holds notable transparency. Moreover, the proposed regularization term is *model-agnostic* and can be flexibly incorporated into the loss function of any type of model architecture.
- We propose to use an *interactive* weight coefficient for both the accuracy loss and fairness regularization terms. This weight coefficient is tuned simultaneously with other key hyperparameters of an AI model (e.g., number of hidden layers, number of hidden neurons, and learning rate of a multiple-layer perception model). Therefore, the fairness-aware travel demand forecasting models can optimally improve fairness while preserving prediction accuracy.

The remaining paper is structured as follows: Section II reviews the related studies. Section III introduces the fairness definitions, metrics and unfairness correction method. We introduce the empirical case studies in Section IV. The modeling results are presented in Section V. Section VI discusses the merits of the proposed methodology, echoes the critical findings, proposes some policy implications and lists several future research directions. Finally, Section VII concludes our study.

II. Literature Review

A. AI Fairness Issues

In recent years, AI methods have been deployed in a broad array of real-world applications due to their strength in producing highly-accurate predictions. However, there has been a growing recognition that, despite predictive superiority, AI and machine learning techniques have also been accompanied by increasing concerns of fairness [7]. Studies from multiple fields have reported that AI algorithms could be discriminatory to the disadvantaged population groups under various applications, including healthcare, criminal justice, credit assessment, translation, among many others [6], [7], [8], [9], [10], [26]. For example, healthcare systems could underestimate the health condition of black patients than white patients, even if they have the same health risk score [10]. If these inherent biases are not addressed, using these AI systems to assist decision-making will worsen the existing social disparities [27].

1) Taxonomy of Fairness Notions: Numerous fairness notions and corresponding mathematical formulations have been proposed for different downstream learning tasks [27]. These fairness notions span various dimensions, including classification vs. regression, group vs. individual and disparate treatment [28]. In classification, multiple fairness notions are created to mitigate "disparate impact", i.e., if practices or policies have disproportionately adverse effects on different groups [29]. For example, *statistical parity* [30], *equality of odds* and and *equality of opportunity* [31]. In regression, notions like *individual/region-based fairness gap* [12], *cross-pair* loss [28] and *equal means* [32] are introduced to address real-world regression applications that require fairness concerns. Fairness notions also branch into the axis of individual and group. Individual fairness requires similar individuals to be treated similarly, while group fairness equalizes the outcome among all groups [30]. Another branch to classify fairness notions is determining whether the *disparate treatment* is allowed. Disparate treatment measures fairness through treatment rather than the outcomes. It addresses both formal classification and intentional discrimination [29], and includes notions like *counterfactual fairness* [33] and *fairness through unawareness* [30]. These fairness notions have laid a solid foundation for defining and measuring fairness in real-world problems.

2) Correcting Unfairness for Multiple Protected Attributes: There are three possible ways to achieve the aforementioned fairness, i.e., correcting the unfairness. First, *pre-processing* the data (e.g., resampling or reweighting) and remove bias before training the models (e.g., [34], [35]). Second, *in-processing*: modifying the algorithms such as including fairness penalty in the loss function [12], [28] or incorporating constraints [36]. Third, *post-processing*: correcting unfairness by adjusting the learned algorithms [31], [37]. In this study, we selected the in-processing techniques due to their transparency (i.e., directly taking fairness into model optimization) and strong capabilities in achieving fairness even when confronted with biased data [38] and the effectiveness in mitigating bias amplification problems (i.e., the trained models amplify the biases in the training data) [39].

In-processing methods involve two categories: implicit method and explicit method [24]. Implicit methods debias the models by implicitly removing bias from the latent representations. They usually hypothesize that if the latent

representations are less biased, the predictions produced from the representations could also be less biased. The implicit methods are commonly used in adversarial learning [40], [41], [42], contrastive learning [43], etc. However, these methods (1) are usually less transparent since we can hardly interpret how the produced latent representations mitigate (or even remove) the unfairness [44], [45] and (2) usually have specific model architectures [42]. Explicit methods focus on explicitly modifying the objective function while keeping the model structure intact, for example, adding fairness-related regularization terms or constraints. Therefore, the explicit methods usually afford greater flexibility and can be applied to a wide range of models. Existing explicit methods include absolute correlation regularization term [5], pairwise fairness loss [28], equal means [32], etc. This study adopts the explicit method by integrating a fairness-related regularization term into the loss function to jointly account for accuracy and fairness.

Achieving multi-attribute fairness has long been an enduring challenge in using in-processing techniques to mitigate unfairness [24]. To date, most of the existing literature purely focused on correcting the unfairness of a single protected attribute [28], [36], [40], [46], [47]. However, mitigating the unfairness of one attribute may increase the unfairness of another attribute [13]. This unexpected outcome may confuse the end-users (e.g., travel demand modelers) and thus hinder the adoption of the fairness-aware models. To tackle this issue, [12] proposed to explicitly correct the unfairness of multiple attributes by simply adding multiple regularization terms (one for each attribute with a corresponding weight) into the loss function. However, when the protected attributes are correlated with each other (which is the case for most travel demand forecasting problems), it could be challenging to determine the appropriate weight for each protected attribute in order to achieve the optimal solution that minimizes the unfairness for the combination of the selected protected attributes. Other related methods include learning fair graph embeddings via adversarial learning [25], disentangled representation learning [48], adding fairness constraints for each protected attribute and achieving fairness via constrained optimization [49], [50]. However, as we discussed, these methods are often less transparent and come with specific model architectures, which hinder their adaptability. As of now, there is a pressing need to develop transparent, effective and flexible methods that can simultaneously account for fairness for multiple protected attributes and can be applied to any model class.

B. Addressing AI Fairness Issues in Travel Demand Prediction

The standard method of estimating travel demand is the four-step model, including trip generation, trip distribution, mode split and traffic assignments [51]. Accordingly, recent studies have started to examine and address the fairness concerns of travel demand forecasting problems spanning across these steps [12], [19], [21], [42], [52], [53]. Specifically, several studies focused on resolving unfairness issues for trip generation forecasting [12], [14], [21], [54]. For example, [12] treated fairness as equal mean per capita travel demand across groups over a period of time and evaluated the fairness issues of several AI methods on demand prediction for ridesourcing services and bike-share systems. Results showed that machine learning spontaneously underestimated the travel demand of disadvantaged people. They also proposed two fairness regularization terms and a corresponding fairness-aware demand prediction model to correct the unfairness. [21] proposed a socially-Equitable Interactive Graph information fusion-based mobility flow prediction system for Dockless E-scooter Sharing (EIGDES) along with a novel regularization term to ensure both the dockless e-scooter prediction accuracy and spatial fairness. The proposed model and the regularize jointly work to penalize demand overestimations and reduce output disparities. In addition, one study [53] has explored addressing fairness issues in trip distribution prediction. Specifically, the authors predicted the Origin-Destination (OD) travel demand by using Multi-Objective Reinforcement Learning (MORL), where the objectives are optimizing transportation network's efficiency and mitigating the demand disparities among different population groups. Certain studies also explored the unfairness issue in travel mode split problems [13], [19]. For example, [13] studied the prediction disparities among population groups by using both a binary logistic regression and a three-layer deep neural network (DNN). They also developed an absolute correlation term as fairness regularizers to mitigate the mode-choice prediction disparities among different population groups. Fairness concerns were also addressed in traffic assignment problems [55], [56], where the objective is usually to obtain optimal travel flows that minimize the discrepancy of user travel time sharing the same link. In addition to four-step model, researchers also developed unfairness correction methods for other tasks, such as traffic safety [11] and infrastructure planning [57]. For example, [11] proposed a Synthetic Minority Oversampling Technique (SMOTE) with the attentive interpretable (TabNet) model to enhance the fairness of traffic crash prediction. These fairness-enhancing methods offer transportation professionals new insights on transportation resource allocations and a novel instrument for designing a fairer transportation ecosystem.

However, there are still two critical knowledge gaps that have yet to be addressed. Firstly, prior research has primarily concentrated on equalizing per capita travel demand among different population groups, but we should note that travel demand disparities may have already been introduced during the data creation process, which is often beyond our control [13], [58]. For example, multiple studies found that rich people are more likely to use ridesourcing services than the poor [59]. That means this behavioral bias among different population groups may naturally exist [60]. However, to date, no study has investigated how to appropriately account for this type of bias, especially for travel demand forecasting models. Second, the existing fairness-aware travel demand forecasting methods necessitate particular model structures, which has very limited adaptability. Thus, developing a model-agnostic (i.e., independent of the model structure) method that can be flexibly adopted by different types of AI models is promising. To date, however, a systematic method in a model-agnostic manner to address fairness issues,

TABLE I
A LIST OF SYMBOLS AND NOTATIONS

Notations	Description	Notations	Description
Indices and Sets		L	overall loss function
G	graph	l	primary loss function for forecasting model
V	the set of nodes	*Variables*	
T	the set of time	$d_{i,j}$	the distance between node i and node j
$\mathcal{J}$	the index set of attributes	z_j^i	the value of the protected attribute j at node i
$\mathcal{I}$	the index set of nodes	$\boldsymbol{z}_j$	the protected attribute j, $\boldsymbol{z}_j = (z_j^i, i \in \mathcal{I})$
$\mathcal{I}_j^+$	the set of advantaged node index	x_t^i	travel demand at node i at time t
$\mathcal{I}_j^-$	the set of disadvantaged node index	$x_t^{i\prime}$	$x_t^{i\prime} = \max\left(x_t^i, 1\right)$
t	timestamp	$\widehat{x}_t^i$	estimated travel demand of node i for time t
E	the set of edges representing the connectivity between two nodes	$\widehat{\boldsymbol{x}}_t$	estimated travel demand at time t, $\widehat{\boldsymbol{x}}_t = (\widehat{x}_t^i, i \in \mathcal{I})$
$w_{i,j}$	the element in the weighted adjacency matrix	$\boldsymbol{x}_t$	ground truth travel demand at time t, $\boldsymbol{x}_t = (x_t^i, i \in \mathcal{I})$
$\boldsymbol{W}$	weighted adjacency matrix	$\widehat{\boldsymbol{Y}}_t$	ground truth travel demand of next M time intervals starting from t, $\widehat{\boldsymbol{Y}}_t = [\widehat{\boldsymbol{x}}_t, \cdots, \widehat{\boldsymbol{x}}_{t+M-1}]$
N_j^+	the size of the set of advantaged node index	$\boldsymbol{Y}_t$	ground truth travel demand of next M time intervals starting from t, $\widehat{\boldsymbol{Y}}_t = [\widehat{\boldsymbol{x}}_t, \cdots, \widehat{\boldsymbol{x}}_{t+M-1}]$
N_j^-	the size of the set of disadvantaged node index	$\boldsymbol{X}_t$	input historical K travel demand before time t, $\boldsymbol{X}_t = [\boldsymbol{x}_{t-K}, \cdots, \boldsymbol{x}_{t-1}]$
Parameters		$\boldsymbol{Z}$	the matrix of protected attributes, $\boldsymbol{Z} = [\boldsymbol{z}_j, j \in \mathcal{J}]$
K	length of input historical sequence	p_j^i	the binary indicator indicating if node i is belonging to advantaged ($p_j^i = 1$) or disadvantaged ($p_j^i = 0$) groups for protected attribute j
M	length of output sequence	$\boldsymbol{e}_t$	the prediction accuracy at time t, $\boldsymbol{e}_t = (e_t^i, i \in \mathcal{I})$
N	the number of nodes	e_t^i	the prediction accuracy of node i at time t
λ	interactive weight coefficient	$r(\boldsymbol{e}_t, \boldsymbol{z}_j)$	the correlation between prediction accuracy at time t and the protected attribute z_j
Q	the total number of protected attributes	$R(\boldsymbol{e}_t, \mathbf{Z})$	multiple correlation coefficient between prediction accuracy at time t and a set of protected attributes
Functions		$\bar{\boldsymbol{e}}_t$	the expectation of prediction accuracy $\boldsymbol{e}_t$
$h(\cdot)$	function of the travel demand forecasting problem	$\bar{\boldsymbol{z}}_j$	the expectation of protected attribute $\boldsymbol{z}_j$

especially for travel demand forecasting problems, is still lacking.

III. METHODOLOGY

The methodological framework is outlined as follows. The travel demand forecasting problem will be mathematically defined in Section III-A. In Section III-B, we will introduce the fairness metrics used in the proposed methodology, followed by the unfairness correction approach for multiple attributes (in Section III-C). The notations are summarized in Table I.

A. Travel Demand Forecasting Problem

The goal of travel demand forecasting is to predict the future travel demand for each area (or other spatial unit such as traffic segments) given previously observed time-series data. This study considers travel demand forecasting as a trip generation modeling problem. Specifically, we consider the transportation network as a weighted directed graph $G = (V, E, \boldsymbol{W})$, where V is a set of nodes (i.e., areas or traffic segments) with $|V| = N$; E is a set of edges representing the connectivity between two nodes; and $\boldsymbol{W} \in \mathbb{R}^{N \times N}$ is a weighted adjacency matrix representing the node's proximity (e.g., distance or functional similarity). Given weighted directed graph G with N nodes, we assume time $t \in T$ is a discrete variable where T is a set containing all possible timestamps, let $\boldsymbol{x}_t = (x_t^i, i \in \mathcal{I})$ represent travel demand at time t, where $\mathcal{I}$ is the index set of nodes, x_t^i is the travel demand corresponding to node $i \in \mathcal{I}$ at time t, and let $\boldsymbol{X}_t = \left[\boldsymbol{x}_{t-K}, \cdots, \boldsymbol{x}_{t-1}\right]$ be historical K travel demand before $\boldsymbol{x}_t$. The travel demand forecasting problem could be formulated as learning a function $h(\cdot) : \mathbb{R}^{N \times K} \rightarrow \mathbb{R}^{N \times M}$ which maps the historical K travel demand to travel demand at next M time interval for all nodes in a given graph G. Let $\widehat{\boldsymbol{Y}}_t = [\widehat{\boldsymbol{x}}_t, \cdots, \widehat{\boldsymbol{x}}_{t+M-1}]$ denote the predicted travel demand for next M time interval starting from timestamp t, where $\widehat{\boldsymbol{x}}_t = (\widehat{x}_t^i, i \in \mathcal{I})$ refers to the predicted travel demand at timestamp t for all nodes, then we can mathematically write:

$$h\left(\boldsymbol{X}_t | \, G\right) = \widehat{\boldsymbol{Y}}_t = [\widehat{\boldsymbol{x}}_t, \cdots, \widehat{\boldsymbol{x}}_{t+M-1}] \tag{1}$$

B. Fairness in Travel Demand Forecasting Models

This study defines ***Fairness*** as the ***equality of prediction accuracy***. Intuitively, we assume that the travel demand prediction accuracy should be independent of the protected attributes. Taking racial composition as an example, equality of prediction accuracy suggests that the prediction accuracy for any racial group should be equal.

In this study, we use the Absolute Percentage Error (APE) to measure the predictive accuracy for each node instead of the Mean Absolute Error (MAE) or Root Mean Square Error (RMSE). We believe the magnitude of the travel demand (especially for the emerging mobility) for an advantaged community (e.g., high-income community) should be naturally greater than a disadvantaged community [61]. This type of behavioral bias may largely be introduced during the data creation process instead of applying the algorithm [58], [60]. If we quantify the equality of prediction accuracy by using MAE and RMSE, which are *scale-dependent* [62], [63] (i.e.,

sensitive to the magnitude of the forecasting outcome), the results could be biased and may not accurately reflect the performance disparities across different communities. Instead, APE scales the magnitude and describes the performance by percentage, and is thus *scale-independent*. We believe using APE as a performance metric for prediction accuracy can help cancel out the behavioral bias that has already been embedded in the data.

Recall from the previous section, a travel demand forecasting model is to learn a function h which takes K historical travel demands $\left[\boldsymbol{x}_{t-K}, \cdots, \boldsymbol{x}_{t-1}\right]$ as input and predict travel demand from next M time interval starting from time t, i.e., $\boldsymbol{Y}_t$. We define $\boldsymbol{e}_t = (e_t^i, i \in \mathcal{I})$ to indicate the prediction accuracy (i.e., APE) at time t, and e_t^i is the prediction accuracy of node i at time t. Specifically,

$$e_t^i = \left| \frac{x_t^i - \hat{x}_t^i}{x_t^{i\prime}} \right| \tag{2}$$

where $x_t^i, \widehat{x}_t^i$ are the ground truth and predicted value of node i at time t, respectively; $x_t^{i\prime} = \max\left(x_t^i, 1\right)$ is used to ensure the fraction is defined [64], [65]; e_t^i is the absolute percentage error for node i at time t. The lower the value of e_t^i, the better the predictive performance.

Suppose $\boldsymbol{Z} = [z_j, j \in \mathcal{J}]$ is the matrix of protected attributes of interest, where $\mathcal{J} = \{1, 2, \ldots, Q\}$ is the index set of attributes, where Q is the total number of protected attributes; $\boldsymbol{z}_j = (z_j^i, i \in \mathcal{I})$ represents the protected attribute j, and z_j^i denotes the protected attribute j at node i, $\mathcal{I}$ is the set of index for nodes. Denote p_j^i as a binary indicator indicating if node i is belonging to advantaged (i.e., $p_j^i = 1$) or disadvantaged (i.e., $p_j^i = 0$) groups for protected attribute j, and accordingly let $\mathcal{I}_j^+ = \{i : p_j^i = 1\}$ and $\mathcal{I}_j^- = \{i : p_j^i = 0\}$ represent the set of advantaged and disadvantaged node index for demographic attribute j with size $N_j^+ = |\mathcal{I}_j^+|$ and $N_j^- = |\mathcal{I}_j^-|$, respectively. We note that assigning value for p_j^i, i.e., determining whether each node should be labeled as advantaged or disadvantaged, is case-specific and should be decided by end-users such as the city government, the transportation system operators or the travel demand modelers. Subsequently, ***Equality of Prediction Accuracy*** is defined as:

$$\mathbb{E}\left(\boldsymbol{e}_t | p_j^i = 1\right) = \mathbb{E}\left(\boldsymbol{e}_t | p_j^i = 0\right) \quad \forall j \in \mathcal{J}, \forall i \in \mathcal{I} \tag{3}$$

$$i.e. \quad \frac{1}{N_j^+} \cdot \sum_{i \in \mathcal{I}_j^+} e_t^i = \frac{1}{N_j^-} \cdot \sum_{i \in \mathcal{I}_j^-} e_t^i \quad \forall j \in \mathcal{J} \tag{4}$$

where $\mathbb{E}\left(\boldsymbol{e}_t | p_j^i = 1\right)$ and $\mathbb{E}\left(\boldsymbol{e}_t | p_j^i = 0\right)$ are the conditional expectation of prediction accuracy $\boldsymbol{e}_t$ given $p_j^i = 1$ and $p_j^i = 0$, and represents the mean APE for advantaged group and disadvantaged group respectively. That is, for any protected attribute j, a fair model should have equal prediction accuracy for different groups. Moreover, when a forecasting model is conducted, we could measure the model fairness by quantifying prediction accuracy disparities, especially between nodes with different labels, for instance, low-income communities and high-income communities.

In this study, we introduce ***Prediction Accuracy Gap*** (**PAG**) as a fairness metric to measure prediction accuracy disparity and if fairness/unfairness achieves/occurs. Define:

$$PAG_j = \mathbb{E}\left(\boldsymbol{e}_t | p_j^i = 0\right) - \mathbb{E}\left(\boldsymbol{e}_t | p_j^i = 1\right), \quad \forall i \in \mathcal{I} \tag{5}$$

Intuitively speaking, PAG directly measures the prediction accuracy disparity between these two types of nodes. A high value of PAG indicates that the machine learning model delivers inconsistent predictive performance among nodes; in most cases, the performance is worse in disadvantaged nodes. PAG is also connected with the popular recognition that a model is considered fairer if it suggests a smaller difference between two group's prediction accuracy [66].

In this study, we also use ***Correlation Coefficient*** as another fairness metric. The correlation coefficient can naturally measure the extent to which the predictions are biased on specific protected groups. Intuitively, if fairness is achieved, correlation between prediction accuracy and any protected attribute should be zero. By using correlation coefficient as a measure of fairness, we assume that the target variable (i.e., prediction accuracy) is linearly correlated with the independent variable (i.e., protected attribute).

Recall from the discussions above, $\boldsymbol{e}_t$ is the prediction accuracy (APE) at time t, and $\boldsymbol{z}_t$ refers to the protected attribute j for all nodes. Then, the correlation between prediction accuracy $\boldsymbol{e}_t$ and the protected attribute $\boldsymbol{z}_j$ across all nodes is denoted by $r(\boldsymbol{e}_t, \boldsymbol{z}_j)$. Define:

$$r(\boldsymbol{e}_t, \boldsymbol{z}_j) = \frac{\sum_{i \in \mathcal{I}} \left(e_t^i - \bar{\boldsymbol{e}}_t\right)\left(z_j^i - \bar{\boldsymbol{z}}_j\right)}{\left(\sqrt{\sum_{i \in \mathcal{I}} \left(e_t^i - \bar{\boldsymbol{e}}_t\right)^2}\right)\left(\sqrt{\sum_{i \in \mathcal{I}} \left(z_j^i - \bar{\boldsymbol{z}}_j\right)^2}\right)} \tag{6}$$

where $\bar{\boldsymbol{e}}_t = \mathbb{E}(\boldsymbol{e}_t)$ and $\bar{\boldsymbol{z}}_j = \mathbb{E}(\boldsymbol{z}_j)$. In our experiment, we add small $\epsilon = e^{-20}$ to denominator to keep it always positive. Although correlation coefficient does not require a label for each region, we cannot directly read the prediction accuracy disparity from it.

C. Unfairness Correction Method for Travel Demand Forecasting Models

In this study, we introduce an absolute correlation regularization approach, which adapts the efforts from [5], to mitigate the prediction accuracy disparities existing among groups. In [5], the authors applied this approach to a classification problem by minimizing the false positive rate (FPR) gap between groups. We generalize this approach to a regression setting (i.e., travel demand forecasting problem) by minimizing the prediction accuracy disparities among different communities.

More importantly, including [5], most previous studies have primarily focused on correcting the unfairness of one single attribute. In real-world dataset, however, the debiased model and results could differ among various protected attributes. Also, a model that is fair for one protected attribute could still be unfair for other attributes [13], [24]. One feasible

solution to solve this issue is to consider multiple attributes at the same time when correcting the unfairness of the models. We expected that a fair model should produce fair predictions for all types of attributes instead of focusing solely on one.

Therefore, we propose a methodology that can correct the unfairness for multiple protected attributes. More specifically, we propose to use the ***Multiple Correlation Coefficient*** [67], denoted as R, to measure the correlation between the target variable, i.e., prediction accuracy, and a set of protected attributes (including race, education, age and income). A larger R suggests that a stronger dependence may exist between the target variable and the explanatory variables. We expect that a fair prediction should lead to $R = 0$, or at least, a small value. Accordingly, we will use R as the regularization term in the loss function to account for fairness loss. We should note that the linear model may encounter potential multicollinearity concerns. However, there is no need to address them since the goal of the linear model is forecasting rather than estimating the coefficients [68].

Recall from previous subsections, we will use the prediction accuracy $\boldsymbol{e}_t$ as the target variable and $\boldsymbol{Z} = [z_j, j \in \mathcal{J}]$ to represent the matrix of multiple protected attributes of interest. And, we use $r(\boldsymbol{e}_t, z_j)$ to indicate the correlation between prediction accuracy $\boldsymbol{e}_t$ and the protected attribute z_j across all nodes. Given these notations, we will naturally write the vector of correlations between each protected attribute z_j and prediction accuracy $\boldsymbol{e}_j$, i.e., $\boldsymbol{c} = \left(r(\boldsymbol{e}_t, z_1), r(\boldsymbol{e}_t, z_2), \ldots, r(\boldsymbol{e}_t, z_Q)\right)^\top$, and the correlation matrix calculated by the correlation coefficient among each pair of protected attributes, denoted as Ω, i.e.,

$$\Omega = \begin{pmatrix} r(z_1, z_1) & r(z_1, z_2) & \ldots & r(z_1, z_Q) \\ r(z_2, z_1) & \ddots & & \vdots \\ \vdots & & \ddots & \\ r(z_Q, z_1) & \ldots & & r(z_Q, z_Q) \end{pmatrix}$$

Consequently, the absolute value of the multiple correlation coefficient between $\boldsymbol{e}_t$ and $\mathbf{Z}$), i.e., $R(\boldsymbol{e}_t, \mathbf{Z})$, which is the square root of the coefficient of determination (i.e., R^2) of the linear model [69], can be written as:

$$R(\boldsymbol{e}_t, \mathbf{Z}) = \left| \sqrt{\boldsymbol{c}^\top \Omega^{-1} \boldsymbol{c}} \right|, \tag{7}$$

where $\boldsymbol{c}^\top$ is the transpose of $\boldsymbol{c}$ and Ω^{-1} is the inverse matrix of Ω.

Accordingly, given graph G and a forecasting model $\widehat{\boldsymbol{Y}}_t = h(\boldsymbol{X}_t|G)$, we add the multiple correlation coefficient, R, into the loss function, denoted as $L(\boldsymbol{X}_t, \mathbf{Z}|G)$ as shown in Eq. 8. In this way, the model will simultaneously account for the unfairness issues sourcing from multiple protected attributes. Let $\boldsymbol{Y}_t = [\boldsymbol{x}_t, \cdots, \boldsymbol{x}_{t+M-1}]$ denote the ground truth travel demand of next M time intervals starting from t, mathematically, the loss function of the forecasting model to be minimized, i.e., $L(\boldsymbol{X}_t, \mathbf{Z}|G)$, is written as:

$$L(\boldsymbol{X}_t, \mathbf{Z}|G) = \sum_t \left\{(1-\lambda) \cdot l\left(\boldsymbol{Y}_t, h(\boldsymbol{X}_t|G)\right) + \lambda R(e_t, \mathbf{Z})\right\}, \tag{8}$$

and,

$$l\left(\boldsymbol{Y}_t, h(\boldsymbol{X}_t|G)\right) = \frac{1}{N} \sum_{i=1}^{N} \left(x_t^i - \widehat{x}_t^i\right)^2 \tag{9}$$

In the above equations, x_t^i, $\widehat{x}_t^i$ refer to the ground truth and predicted travel demand for node i at time t, respectively; l is the primary loss function for forecasting model, and in this study, we use mean squared error (MSE) for l; λ is the *interactive* weight coefficient, a concept borrowed from the traditional multi-task learning framework [70], [71], which controls the weight between the prediction loss and the fairness loss.[1] When $\lambda = 0$, the model will be unaware of the fairness; and when $\lambda = 1$, the model will completely focus on correcting the unfairness. We can directly treat λ as a hyperparameter to find the optimal model that effectively addresses fairness while preserving accuracy. The prediction accuracy disparity is captured and mitigated by the correlation regularization term, in Eq. (7). The regularization term is dedicated to shrinking the potential prediction accuracy disparity that existed among groups toward zero. Incorporating it into the loss function enables the machine learning model to automatically keep track of the fairness during training.

Note that when there is only one single protected attribute of interest, the multiple correlation coefficient, i.e., Eq. 7 reduces to Eq. 6.

IV. Case Study

In this section, we will describe two real-world ridesourcing-trip datasets and seven commonly-used travel demand forecasting models used for case studies. Section IV-A and Section IV-B present the data collection and processing process. Table II presents the descriptive statistics of all input variables. In Appendix.A, Fig. 1 displays the spatial distribution of the average ridesourcing demand per hour. We will briefly introduce the selected deep learning and statistical models for unfairness detection and correction in Section IV-C.

A. Chicago Ridesourcing-Trip Data

In this study, we collected the publicly available ridesourcing-trip data from Chicago Data Portal[2] for case study. The data are from November 1, 2018 to March 31, 2019, containing 45,338,599 trips. There are plenty of attributes included in this dataset, but only pick-up locations and timestamps are considered for this research. Since we focused on trip generation (i.e., origin demand) forecasting, all trips are aggregated at the census-tract level and hourly counted. We prepared the data for modeling in the same way as previous studies [59], to account for the missing-data issues and outliers. The data preparation process produced the trip generation data for 711 census tracts. We split the first 70%

[1] In traditional multi-task learning setting, the objective function is usually written as $\lambda * f_1 + (1-\lambda) * f_2$ where f_1 and f_2 are two learning tasks and $\lambda \in [0, 1]$ is the weight on task 1. Here, in our study, fairness is explicitly designed as an additional task alongside prediction accuracy.

[2] https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips-2018-2022-/m6dm-c72p/explore

TABLE II
DESCRIPTIVE STATISTICS

	Chicago				Austin			
	Min	Max	Mean	St. Dev.	Min	Max	Mean	St. Dev.
Target Variable								
Hourly ridesourcing trip demand	0.00	2150.00	12.01	41.26	0.00	820.00	1.41	8.81
Demographic Characteristics								
Race: Percentage of white population	0.00	0.97	0.47	0.32	0.44	0.99	0.76	0.13
Edu: Percentage of population with a bachelor's degree or above	0.01	0.95	0.37	0.29	0.06	0.92	0.49	0.26
Age: Percentage of young population (aged 18 - 44)	0.21	0.89	0.44	0.12	0.14	0.98	0.48	0.14
Income: Percentage of low-income households	0.01	0.79	0.27	0.16	0.01	0.84	0.18	0.12

data for training, the following 10% for validation and the remaining for testing. The census-tract-level demographic data (i.e., protected attributes) were collected from the American Community Survey (ACS) 2015-2019 5-year estimates data, including the percentage of white, the percentage of low-income households, the percentage of population with a bachelor's degree or above and the percentage of young populations (with age in 18-44).

B. Austin Ridesourcing-Trip Dataset

This study also collected ridesourcing-trip data from RideAustin[3] for case study. The data ranges from October 1, 2016 to April 13, 2017, including 1,259,574 trips in total. Similar to the case study in Chicago, we only retained pick-up locations and the corresponding timestamps from the dataset for empirical analysis. All ridesourcing trips were aggregated at the census-tract level on an hourly basis. Finally, the prepared dataset includes 191 census tracts. The first 70% of the whole dataset was split for model training, followed by the following 10% for validation and 20% for testing. Four protected attributes, including the percentage of white, the percentage of low-income households, the percentage of population with a bachelor's degree or above and the percentage of young populations (aged 18-44) were also collected from ACS 2013-2017 5-year estimates data.

C. Model Comparison

In this study, we applied seven models as the major baseline models to measure the fairness metrics and perform the bias mitigation. We also compared their performance with historical average method. All used models are detailed as follows:

- **Historical Average** (HA): We calculate the historical average travel demand using the mean values of all observations from the inputted sequence.
- **Multivariate Linear Regression** (MLR): MLR is frequently used in machine learning studies as the benchmark model. This study treats observations at every timestamp t as a covariate.
- **Autoregressive Integrated Moving Average Model** (ARIMA): ARIMA is one of the most fundamental statistical models for forecasting time-series data [72]. ARIMA consists of three basic parts: auto-regressive, first-differencing and moving-average part. The order of the auto-regressive (p) and moving-average (q) and the degree of first-differencing (d) included should be prespecified before building the model. In this study, we established ARIMA model to predict the travel demand for all areas at once.
- **Multiple Layer Perception** (MLP): MLP is a commonly-used deep neural net model. In this study, the model architecture is set as 1 hidden layer with 300 hidden linear neurons. A drop-out layer rate 0.01 is set after the hidden layer to avoid overfitting.
- **Gated Recurrent Unit** (GRU): GRU is a widely-adopted Recurrent Neural Network (RNN) model with gated hidden neurons [73]. GRU can generate the predicted travel demand $x_{i,t+1}$ by inputting the hidden status at timesampe $t-1$ and the travel demand at timestamp $x_{i,t}$. In this way, GRU can dynamically capture the travel demand information at the current timestamp while maintaining the historical demand trend. We use GRU model for forecasting the travel demand for all nodes at once.
- **Temporal Graph Convolution Network** (T-GCN): T-GCN can capture the spatial dependency and temporal information at the same time [74]. Specifically, the spatial dependency is calibrated by the spatial adjacency graph G_{adj}, where 1 indicates two nodes are spatially adjacent and 0 otherwise. T-GCN takes the hidden status at timestamp $t-1$ and the graph-convolution-processed travel demand information at timestamp t as the input. Therefore, T-GCN can effectively deal with data that have strong spatial dependency such traffic speed data.
- **Convolutional Long-short Term Memory (ConvLSTM)**: ConvLSTM is one of the most novel approaches for spatio-temporal forecasting problem [75]. ConvLSTM has a convolution structure in both the input-to-state and state-to-state transitions; it determines a certain cell's future states by considering the inputs and past states from its local neighbors. This characteristic allows it a more powerful strength in handling spatio-temporal correlations. In this study, the convolutional kernel size of the ConvLSTM is set to 5.
- **Spatio-Temporal Graph Convolution Network (STGCN)**: STGCN is an effective approach for spatio-temporal traffic flow forecasting [76]. STGCN consists of several spatio-temporal convolution (ST-Conv)

[3]https://data.world/ride-austin/ride-austin-june-6-april-13

blocks. Each block has a "sandwich"-like structure: two gated sequential convolution layers and one spatial graph convolution layer in between. This allows STGCN to distill the most useful spatial features and capture the most essential temporal features collectively. In this study, we set the number of ST-Conv blocks as 2. Let $d_{i,j}$ denote the distance between node i and node j, the element in the weighted adjacency matrix, i.e., $w_{i,j} \in \boldsymbol{W}$, is given by:

$$w_{i,j} = \begin{cases} \exp\left(-\frac{d_{ij}^2}{\sigma^2}\right), & i \neq j \text{ and } \exp\left(-\frac{d_{ij}^2}{\sigma^2}\right) \geq \alpha \\ 0, & \text{otherwise.} \end{cases}, \tag{10}$$

where σ^2 and α, assigned as 10^4 and 0.5, are thresholds that control the sparsity of $\boldsymbol{W}$.

MAE, RMSE and Mean Absolute Percentage Error (MAPE) are employed to evaluate the model's prediction accuracy. Note that we only calculate MAPE for samples with ridesourcing demand larger than 10 since MAPE is sensitive to small values [3], [77]. Correlation coefficient and PAG are used to detect and evaluate the model's prediction fairness.

V. Results

This section sequentially reports the modeling results of all benchmark models, the evaluations of their underlying fairness issues and the results after applying our proposed unfairness correction approach. We conducted empirical experiments using the real-world ridesourcing-trip data in Chicago, IL and Austin, TX. The analytical spatial unit is census tract. We incorporate the regularization term into the loss function for all models. All experiments were completed in a Pytorch environment using an Ampere A-100 GPU. We tuned the hyperparameters such as batch size and sequence length under each fairness weight λ using grid search. We built our models with Adam optimizer [78]. Early stopping method is also taken to avoid overfitting problems. In this study, we use 60 and 40 percentile statistics for the protected attributes as the threshold to determine the label (i.e., p_j^i) of each node (e.g., census tract). For instance, the 60 percentile of white population percentage attribute is 62.35%, for nodes with white population percentage over 62.35% are labeled as advantaged. We also test the sensitivity of such threshold settings by using 50 and 50, 70 and 30 percentile statistics. The detailed results are presented in Supplementary Materials. Overall, major findings hold across different threshold settings, which further highlights the effectiveness of the proposed unfairness correction method.

A. Unfairness Detection

The predictive performance and two fairness metrics (i.e., correlation [Corr] and prediction accuracy gap [PAG]) of all models with respect to four protected variables are presented in Table III and Table IV.

We show the results of the predictive performance for each benchmark in Chicago ridesourcing-trip data (Table III) and Austin ridesourcing-trip data (Table IV).

Regarding prediction accuracy, all benchmark models show a similar trend across two case studies. The performance ranking is ConvLSTM $\approx$ STGCN $>$ GRU $>$ T-GCN $>$ MLP $>$ ARIMA $>$ HA. It indicates that the prediction accuracy gradually increases as the model becomes more complex. Two convolution models, i.e., STGCN and CovLSTM, are best-performing among all models. Both STGCN and ConvLSTM can incorporate spatial and temporal information through the convolution blocks, which enhance their prediction power. Among two RNN-based models, GRU outperformed T-GCN for both MAE, RMSE and MAPE_{10}. MLP, due to its simple model architecture, underperformed all neural network-based models. Compared with deep neural networks, traditional statistical models, i.e., MLR and ARIMA, have relatively low prediction accuracy. However, their performance still significantly outperformed HA. MLR and ARIMA both have a prespecified (linear) model structure and cannot capture the nonlinearity between the inputs and target variables, which restricts the predictive capability.

Regarding fairness issues, for Chicago ridesourcing-trip data, Table III shows that HA exhibits completely inverse relationships in correlation and gap compared with other models. Since HA has the worst predictive performance, the corresponding fairness metrics could be unreliable. The results illustrate that both statistical and deep learning models have evident fairness issues. Protected attributes, including race, education and age, are negatively correlated with the prediction accuracy which means that communities with high proportion of white population, high education-attainment rate and more young people have high prediction accuracy. Income level is positively related to predictive performance, indicating that communities with more low-income households may have higher perdition error. In terms of magnitude, we found that education and age have the largest value of correlation with prediction accuracy, followed by income and race. Although there are variations in the magnitude of correlations, the signs for all protected attributes among all models except for HA are consistent. In addition to correlations, we also explored the PAG between the advantaged groups and disadvantaged groups. Table III presents that all gaps have a positive value (except for HA), indicating that the prediction error for disadvantaged groups is higher than for advantaged groups. Additionally, the prediction accuracy disparity is more pronounced for education and age than for race and income.

For Austin ridesourcing-trip data, all benchmark models demonstrate a similar performance (both trend and direction of associations) compared with using Chicago dataset. However, results showed that the fairness issues are relatively subdued in Austin dataset. In other words, the extent of unfairness (as shown by correlation coefficient and PAG) is notably diminished in comparison to the Chicago dataset. Notably, Table IV shows that prediction accuracy is less biased regarding race and income. Two best-performing models (i.e., STGCN and ConvLSTM) may produce satisfying fair predictions. For example, the correlation between prediction accuracy and race delivered by ConvLSTM is 0.000 and the PAG regarding race is only -0.391%. This evidence indicates that the unfairness in prediction

TABLE III
MODELING RESULTS OF THE BENCHMARKS IN CHICGAO

Models	MAE	RMSE	$MAPE_{10}$	Race		Edu		Age		Income	
				Corr	PAG (%)	Corr	PAG (%)	Corr	PAG (%)	Corr	PAG (%)
HA	7.703	27.630	0.568	0.062	-33.467	0.072	-54.311	0.040	-47.628	-0.037	-31.359
MLR	5.535	12.973	0.413	-0.047	7.689	-0.081	3.838	-0.095	6.097	0.047	6.458
ARIMA	4.541	12.259	0.341	-0.054	4.494	-0.121	9.589	-0.131	10.969	0.059	4.254
MLP	3.918	10.147	0.303	-0.053	4.025	-0.126	9.182	-0.137	10.737	0.062	4.175
GRU	3.715	9.069	0.274	-0.054	5.579	-0.142	15.820	-0.131	12.127	0.074	6.618
TGCN	4.705	9.993	0.319	-0.088	19.783	-0.146	30.972	-0.151	30.904	0.099	18.393
STGCN	**3.012**	8.539	**0.233**	-0.118	9.007	-0.280	21.185	-0.288	22.723	0.136	8.843
ConvLSTM	3.246	**8.176**	0.256	-0.074	8.978	-0.141	12.474	-0.148	13.986	0.104	9.570

Notes: Corr represents correlation. All correlations are statistically significant at 1% confidence level.

TABLE IV
MODELING RESULTS OF THE BENCHMARKS IN AUSTIN

Models	MAE	RMSE	$MAPE_{10}$	Race		Edu		Age		Income	
				Corr	PAG (%)	Corr	PAG (%)	Corr	PAG (%)	Corr	PAG (%)
HA	1.655	8.538	0.652	-0.030	5.116	-0.106	6.073	-0.008	-8.881	0.026	6.170
MLR	1.324	4.280	0.370	-0.008	0.572	-0.041	4.118	-0.056	6.561	-0.029	-0.784
ARIMA	1.335	4.695	0.392	-0.008	0.316	-0.048	4.940	-0.078	8.578	-0.047	-1.758
MLP	1.297	4.163	0.366	-0.008	0.467	-0.048	5.030	-0.074	8.156	-0.042	-1.073
GRU	1.064	3.654	0.315	-0.044	2.166	-0.136	10.639	-0.110	10.648	-0.026	-0.307
TGCN	1.357	3.911	0.375	0.010	-1.983	-0.049	5.627	-0.098	12.801	-0.087	-6.796
STGCN	**1.042**	4.064	0.369	**-0.034**	**1.234**	-0.178	10.933	-0.179	11.614	**-0.081**	**-1.013**
ConvLSTM	1.057	**3.162**	**0.314**	**0.000**	**-0.391**	-0.088	5.289	-0.080	4.642	**-0.024**	**0.888**

Notes: Corr represents correlation. All correlations are statistically significant at 1% confidence level.

accuracy should be of little concern for this protected attribute.

B. Unfairness Correction

We tuned a set of values of λ (i.e., the weight for fairness loss) by grid search to validate the effectiveness of the proposed unfairness correction method. Table V and Table VI present the results of simultaneously mitigating the unfairness issues for multiple protected attributes across two case studies. We only present the best λ (i.e., the one that can significantly improve fairness while largely preserving prediction accuracy) from the empirical experiments. We also add experimental results of correcting unfairness of only one single attribute at the bottom of each table for comparison. For the sensitivity analysis of λ, please refer to Section V-D. As discussed in previous section, only very limited prediction accuracy disparities are detected on race (percentage of white population) and income (percentage of low-income households) in the case study of Austin (as shown in Table IV). Thus, we decided to only correct the unfairness of prediction accuracy manifested in education (percentage of bachelor's degree holders) and age (percentage of young population) in this case.

There are several key findings to highlight. First, results of the multi-attribute scenario show great consistency across two datasets. Table V and Table VI show that in almost all trails, incorporating a small fairness weight can significantly reduce the absolute value of the correlation and PAG across all protected attributes. For example, in Chicago dataset, incorporating only 0.050 fairness weight for T-GCN can lead to 93.131%, 90.673%, 92.989%, 73.803% reduction of the absolute values of the PAG for race, education, age and income, respectively. In the meantime, the correlation between prediction accuracy and protected attributes also improved more than 75%, but RMSE only increased by 3.535%. In Austin dataset, setting λ as 0.025 for ConvLSTM yields 68.973% and 88.496% PAG shrinkage on education and age by sacrificing only 5.661% and 1.118% increase on RMSE (from 3.162 to 3.341) and $MAPE_{10}$ (from 0.314 to 0.318), respectively.

Second, the effects of the proposed unfairness correction method vary across models and protected attributes. For example, Table V shows that when mitigating the income bias, setting λ as 0.025 only reduces 24.558% of the PAG in absolute value for STGCN; while for ConvLSTM, the same setting can lead to a 81.983% reduction. In addition, the case study on Chicago ridesourcing-trip data reveals that compared with education and age, the absolute value of PAG for race and income are more likely to be reduced by MLP. While ConvLSTM shows more strength in reducing PAG for education and age than race and income.

Third, by choosing an appropriate λ, both fairness and accuracy can be improved at the same time. Taking Austin dataset as an example, adding 0.5 fairness weight on MLP can simultaneously reduce the absolute value of PAG and correlations for all protected attributes while even reducing RMSE and $MAPE_{10}$ by 0.128% and 7.865%, respectively.

Moreover, we found that MLR and ARIMA showed limited capabilities in mitigating unfairness. In Chicago dataset, the prediction accuracy disparities of education and age (as shown in the change of PAG) for MLR and ARIMA even increased after debiasing multiple protected attributes. Also, our examination of the Austin dataset indicated that after incorporating the proposed fairness regularization term, although the PAG

TABLE V
MULTI-ATTRIBUTE UNFAIRNESS CORRECTION IN CHICAGO

Multi-attribute (debiasing four selected attributes)								
Model		MLR	ARIMA	MLP	GRU	T-GCN	STGCN	ConvLSTM
λ		0.025	0.025	0.1	0.075	0.05	0.025	0.025
RMSE		12.984 (-0.088%)	12.308 (-0.397%)	10.169 (-0.215%)	9.516 (-4.932%)	10.346 (-3.535%)	8.569 (-0.349%)	8.390 (-2.622%)
$MAPE_{10}$		0.412 (0.288%)	0.341 (-0.072%)	0.297 (1.946%)	0.275 (-0.278%)	0.318 (0.295%)	0.228 (2.176%)	0.255 (0.671%)
Race	Corr	-0.010 (79%)	-0.005 (89.968%)	-0.026 (51.444%)	-0.015 (72.92%)	-0.011 (87.705%)	0.005 (95.586%)	-0.01 (86.283%)
	PAG(%)	0.453 (94.108%)	-2.227 (50.448%)	0.474 (88.233%)	-3.204 (42.566%)	-1.359 (93.131%)	-5.002 (44.465%)	4.107 (54.251%)
Edu	Corr	-0.011 (86.578%)	-0.011 (90.493%)	-0.03 (76.489%)	-0.031 (78.531%)	-0.035 (75.977%)	0.005 (98.351%)	-0.030 (78.854%)
	PAG(%)	-14.994 (-290.648%)	-11.128 (-16.052%)	-2.978 (67.573%)	-2.093 (86.767%)	-2.889 (90.673%)	-12.368 (41.619%)	-1.680 (86.529%)
Age	Corr	-0.009 (90.262%)	-0.01 (92.076%)	-0.039 (71.762%)	-0.037 (71.407%)	-0.031 (79.606%)	0.005 (98.29%)	-0.035 (76.398%)
	PAG(%)	-12.834 (-110.5%)	-7.948 (27.543%)	-2.418 (77.481%)	2.013 (83.405%)	2.167 (92.989%)	-6.761 (70.247%)	0.552 (96.05%)
Income	Corr	0.012 (73.321%)	0.007 (87.26%)	0.018 (71.219%)	0.016 (78.883%)	0.013 (86.674%)	-0.003 (97.501%)	0.015 (85.988%)
	PAG(%)	-3.042 (52.899%)	-5.295 (-24.478%)	0.003 (99.917%)	-1.106 (83.292%)	-4.819 (73.803%)	-6.671 (24.558%)	-1.724 (81.983%)
Single-attribute (only debiasing Income)								
Income	λ	0.025	0.025	0.05	0.075	0.05	0.025	0.05
	RMSE	12.978 (-0.04%)	12.261 (-0.017%)	10.159 (-0.117%)	9.265 (-2.157%)	10.284 (-2.909%)	8.687 (-1.734%)	8.223 (-0.577%)
	$MAPE_{10}$	0.413 (0.052%)	0.339 (0.633%)	0.295 (2.527%)	0.283 (-3.117%)	0.327 (-2.657%)	0.234 (-0.297%)	0.250 (2.463%)
	Corr	0.004 (91.434%)	0.008 (86.32%)	0.013 (78.507%)	-0.002 (96.888%)	-0.004 (96.39%)	-0.002 (98.488%)	0.004 (96.629%)
	PAG(%)	-2.864 (55.643%)	-3.490 (17.948%)	-0.426 (89.786%)	-5.137 (22.376%)	-9.536 (48.153%)	-6.263 (29.172%)	-3.670 (61.647%)

Notes: Corr represents correlation. PAG refers to prediction accuracy gap. The value inside each bracket refers to the percentage change of metric in absolute value. It is computed as: $(|o| - |m|) * 100\%/|o|$, with o denoting the initial value obtained from the fairness-unaware model and m representing the final value from the fairness-aware model. A positive value indicates the improvement while a negative value indicates the reduction.

for MLR and ARIMA decreased, the magnitude of this reduction was comparatively modest in comparison to other models. In fact, these two models are less flexible compared with other deep learning models since they have a pre-specified model structure. We believe that this inherent limitation could hinder their effectiveness in addressing fairness concerns.

Lastly, in most cases, our proposed multi-attribute unfairness correction method shows better performance in reducing disparities of prediction and preserving accuracy compared with only debiasing a single attribute, especially for complex deep learning models (e.g., MLP, GRU, T-GCN, STGCN and ConvLSTM). For example, Table V shows that when considering multiple attributes together, ConvLSTM can close more than 81% of PAG of income in absolute value; while for the single-attribute scenario, the PAG is only reduced by around 60%. However, we also observed in certain cases, single-attribute unfairness correction could produce fairer performance. For example, GRU is found to be more effective in reducing PAG when only debiasing age for Austin dataset.

C. Comparison Between Multi-Attribute and Single-Attribute Unfairness Correction

To provide a more comprehensive demonstration of the efficacy of the proposed multi-attribute unfairness correction approach and to pinpoint potential shortcomings in the single-attribute bias correction method, we conduct a comparative analysis of unfairness correction outcomes achieved through debiasing the age variable alone versus debiasing multiple attributes simultaneously. We have chosen the top-performing model, i.e., ConvLSTM, for demonstration. The resulting findings can be found in Table VII.

We found that correcting unfairness regarding one attribute might even create more biases for other protected attributes, which aligns with one previous study [13]. This finding

TABLE VI
MULTI-ATTRIBUTE UNFAIRNESS CORRECTION IN AUSTIN

Multi-attribute (debiasing two selected attributes)								
Model		MLR	ARIMA	MLP	GRU	T-GCN	STGCN	ConvLSTM
λ		0.025	0.025	0.5	0.05	0.4	0.05	0.025
RMSE		4.280 (0.005%)	4.697 (-0.037%)	4.158 (0.128%)	3.983 (-9.000%)	3.694 (5.541%)	4.691 (-15.433%)	3.341 (-5.651%)
$MAPE_{10}$		0.370 (-0.037%)	0.393 (-0.327%)	0.337 (7.865%)	0.334 (-5.985%)	0.434 (-15.748%)	0.361 (2.166%)	0.318 (-1.118%)
Edu	Corr	-0.009 (78.071%)	-0.01 (79.285%)	-0.004 (91.677%)	-0.007 (94.849%)	0.002 (95.877%)	-0.002 (98.874%)	0.01 (88.668%)
	PAG(%)	1.938 (52.941%)	2.41 (51.213%)	-0.171 (96.6%)	-0.365 (96.569%)	2.038 (63.783%)	1.401 (87.186%)	-1.641 (68.973%)
Age	Corr	-0.01 (82.214%)	-0.014 (81.956%)	-0.015 (79.841%)	-0.028 (74.578%)	0.000 (100%)	0.025 (86.027%)	-0.006 (92.489%)
	PAG(%)	2.342 (64.305%)	3.458 (59.685%)	0.833 (89.786%)	4.180 (60.745%)	1.045 (91.836%)	-1.269 (89.073%)	-0.534 (88.495%)
Single-attribute (only debiasing Age)								
Age	λ	0.075	0.05	0.1	0.025	0.025	0.05	0.05
	RMSE	4.281 (-0.019%)	4.703 (-0.165%)	4.172 (-0.208%)	4.268 (-16.8%)	3.692 (5.593%)	4.392 (-8.075%)	3.354 (-6.062%)
	$MAPE_{10}$	0.370 (-0.037%)	0.392(-0.071%)	0.366 (-0.063%)	0.343 (-8.841%)	0.375 (-0.013%)	0.363 (1.624%)	0.335 (-6.524%)
	Corr	-0.006 (89.122%)	0.000 (99.553%)	-0.003 (96.292%)	-0.003 (96.292%)	-0.014 (85.273%)	-0.009 (94.822%)	0.000 (99.559%)
	PAG(%)	0.839 (87.212%)	2.961 (65.48%)	-0.081 (99.007%)	0.923 (91.332%)	5.788 (54.783%)	1.331 (88.539%)	0.771 (83.389%)

Notes: Corr represents correlation. PAG refers to prediction accuracy gap. The value inside each bracket refers to the percentage change of metric in absolute value. It is computed as: $(|o| - |m|) * 100\%/|o|$, with o denoting the initial value obtained from the fairness-unaware model and m representing the final value from the fairness-aware model. A positive value indicates the improvement while a negative value indicates the reduction.

TABLE VII
PERFORMANCE COMPARISON BETWEEN ONLY DEBIASING AGE AND SIMULTANEOUSLY DEBIASING MULTIPLE ATTRIBUTES

Model	λ	RMSE	$MAPE_{10}$	Race	Edu	Age	Income
Chicago							
ConvLSTM(Original)	0	8.176	0.256	8.978	12.474	13.986	9.570
ConvLSTM (Single of Age)	0.05	8.317	0.260	**11.111**	1.553	-6.224	**9.716**
ConvLSTM (Multi)	0.025	8.390	0.255	4.107	-1.68	0.552	-1.724
Austin							
ConvLSTM(Original)	0	3.162	0.314	-0.391	5.289	4.642	0.888
ConvLSTM (Single of Age)	0.05	3.354	0.335	**2.027**	3.993	0.771	**8.238**
ConvLSTM (Multi)	0.025	3.341	0.318	1.558	-1.641	-0.534	0.295

highlights the importance of considering multiple protected attributes at once. Specifically, results showed that compared with the original model that purely focused on prediction accuracy, solely correcting unfairness of age variable could indeed help drop the absolute value of PAG. However, by only considering age, the PAG for other variables, especially for race and income variables, even increases. For example, in Austin dataset, debiasing only Age shrank the PAG from 4.642% to 0.771% by significantly sacrificing the PAG of income from 0.888% to 8.238%. This unexpected outcome may further shed light on the fact that the transportation resource allocations intended to be fair for distinct age groups could nonetheless still be unfair regarding communities with different income levels. Notably, the results showed that the proposed multi-attribute unfairness correction method can effectively debias multiple protected attributes and in almost all cases the absolute value of PAG is significantly dropped compared with the original model without sacrificing too much prediction accuracy.

D. Sensitivity Analysis of Fairness Weight

We also explored the influence of the fairness weight, i.e., λ, in shaping the interaction between accuracy and fairness based on the predictive performance of seven models with four protected attributes. Fig. 2 presented in Appendix.C illustrates the sensitivity analysis of λ in determining accuracy and fairness. The x-axis is the value of λ while the y-axis is the performance metrics (RMSE, $MAPE_{10}$, correlation coefficient and PAG). Generally, the accuracy for deep learning models decreases when λ gradually increases. While for traditional statistical models such as MLP and ARIMA, the marginal effect of λ is relatively small. We also found that increasing λ may even help improve the prediction performance for MLP and ARIMA. One possible reason could be that the added fairness regularization term helps reduce the variances of the predictions so that such linear models achieve a better performance on testing data. Figures show that as λ grows, the correlation will first drastically increase/decrease, and then remain flat or slightly change. Notably, setting a small weight

($\lambda \leq 0.1$) can lead the correlation drop to around 0. The PAG shows a decreasing trend as λ gradually increases. But in most cases, the gap may get over-corrected when λ is greater than 0.1. According to the tables shown in Section V-B, a suitable fairness weight possibly exists in the range between 0 to 0.1. This finding further reinforces the effectiveness of our proposed unfairness correction approach: incorporating only a small amount of weight for fairness can lead to a significant improvement in producing fair predictions. We also found that increasing fairness weight may not monotonically reduce the PAG. This finding echoes the results in [13], where they showed that increasing fairness weight might even extend the PAG. Our computational experiments show that this scenario frequently occurs for traditional statistical models. This finding also suggests the need for more fine-grained searching ranges of λ when conducting hyperparameter tuning. Overall, the sensitivity of the effects of λ shows great consistency across two case studies. Finally, we noticed that in Austin case, setting fairness weight as 0.4 for GRU led to a substantial increase in RMSE and PAG. One possible reason could be that this combination of hyperparameters might explode the gradients and thus lead to this numerical instability.

E. Comparison With Benchmark Fairness Regularizers

This study compares the performance of the proposed unfairness correction approach (i.e., the absolute correlation regularization term) with three state-of-the-art benchmark regularizers, including Equal Mean (EM) [32], Region-based Fairness Gap (RFG) and Individual-based Fairness Gap (IFG) [12]. For experiments, we only consider single-attribute scenario as these three benchmark regularizers are explicitly designed for addressing unfairness of a single protected attribute. For Chicago Ridesourcing dataset, we select race (percentage of white population) for model debiasing; while for RideAustin dataset, education variable (percentage of bachelor holders) is chosen for comparison. All benchmark regularizers are set with the best-performing λ yielded by our proposed method for comparison.

Table VIII presents the comparative analysis between our proposed method (i.e., absolute correlation regularizer) and three state-of-the-art benchmark regularizers. Results unequivocally show that the proposed method evidently outperforms other methods in terms of preserving prediction accuracy as well as approximating equality of prediction accuracy. Among all regularizers, EM delivers the worst performance. This is expected since EM focuses on balancing the target variable (i.e., ridesourcing demand) of disadvantaged and advantaged groups instead of the prediction accuracy. However, this method could be questionable since the variations in ridesourcing usage between different population groups may naturally exist due to socioeconomic and demographic disparities [61]. RFG and IFG tend to yield improved outcomes in terms of both accuracy and fairness when compared to EM. Moreover, in certain scenarios, their performance (especially for correlation and RMSE) surpasses that of the proposed method. We attribute this to their capabilities to effectively reduce variations in per capita travel demand for each individual population group, as indicated in [12]. However, these two metrics may still not be able to fully account for the inherent disparities of different population groups in generating travel demand [59]. In most cases, especially for deep learning models with more complex model architectures, the proposed method can significantly help reduce the PAG between disadvantaged and advantaged groups while largely preserving the prediction error. Although in some cases the proposed method may not always be the best-performing one regarding both RMSE and MAPE_{10}, the accuracy still remains satisfactory.

VI. Discussion

The above sections demonstrate the modeling results of our proposed unfairness correction method. In this section, we will discuss the merits of the unfairness correction method, policy implications, and the limitations of the work and future research directions.

A. Merit of the Unfairness Correction Method

The merits of the proposed unfairness correction method are threefold.

First, *a new regularizer to simultaneously debias multiple protected attributes*. The current literature rarely discusses how to effectively address fairness issues for multiple protected attributes. However, designing a method that can accommodate various fairness needs is necessary for real-world applications [24]. This study addresses this issue by proposing to use **Multiple Correlation Coefficient** (i.e., R of a linear model) as a regularization term and incorporating it into the loss function. The multiple correlation coefficient can directly measure the correlation between the target variable (i.e., prediction accuracy) and a set of protected demographic variables (i.e., race, age, education and income). By minimizing the coefficient of multiple correlation, AI models can simultaneously debias multiple protected attributes. Unlike adding multiple regularization terms (one for each attribute) [12], this approach is straightforward and easy to implement, and thus there is no need to fine-tune the fairness weight for different attributes (only one is enough). Also, this approach has little concern about the *multicolinearity* issues among different protected attributes (as shown in Appendix.B), since the goal of the linear model is to use the set of protected attributes to forecast the prediction errors instead of estimating and interpreting the beta coefficients [68]. Overall, our proposed unfairness correction method enables future studies to flexibly debias multiple protected attributes of interests.

Second, *flexibility and transparency*. The proposed unfairness correction method is model-agnostic and may be generalizable for different applications and different data modalities. We implemented the unfairness correction method on both statistical and deep learning models. Results jointly demonstrated that, generally, this approach could mitigate the unfairness while only slightly reducing the overall accuracy. Specifically, we correct the unfairness by incorporating an explicitly designed absolute correlation regularization term into the loss function without modifying the model structure. It allows the unfairness correction method great flexibility to be independent of the underlying model. Scholars can thus

TABLE VIII
COMPARISON WITH STATE-OF-THE-ART BENCHMARK REGULARIZERS

Model	Regularizer	Chicago (Race)					Regularizer	Austin (Edu)				
		λ	RMSE	$MAPE_{10}$	Corr	PAG(%)		Lambda	RMSE	$MAPE_{10}$	Corr	PAG(%)
MLR	$r(e_t, z_j)$	0.025	**12.979**	**0.413**	-0.013	**0.623**	$r(e_t, z_j)$	0.025	4.281	**0.370**	-0.015	**-0.318**
	EM	0.025	15.914	0.470	-0.123	55.906	EM	0.025	**4.270**	0.372	-0.030	2.829
	RFG	0.025	13.095	0.414	**0.001**	-4.809	RFG	0.025	4.274	0.371	-0.055	5.609
	IFG	0.025	13.046	0.413	0.002	-4.637	IFG	0.025	4.279	0.370	**-0.005**	0.621
ARIMA	$r(e_t, z_j)$	0.025	**12.266**	**0.338**	-0.017	**-0.152**	$r(e_t, z_j)$	0.025	4.701	0.392	**0.005**	**0.071**
	EM	0.025	15.770	0.407	-0.134	57.975	EM	0.025	4.702	0.396	-0.047	4.730
	RFG	0.025	12.432	0.342	**0.002**	-5.398	RFG	0.025	4.698	0.394	-0.071	7.268
	IFG	0.025	12.369	0.341	0.007	-7.002	IFG	0.025	**4.696**	**0.392**	-0.013	1.776
MLP	$r(e_t, z_j)$	0.050	**10.170**	0.296	**-0.023**	**0.394**	$r(e_t, z_j)$	0.050	4.166	**0.355**	**-0.012**	**0.774**
	EM	0.050	16.890	0.320	-0.077	17.377	EM	0.050	4.120	0.362	-0.050	4.898
	RFG	0.050	10.228	**0.293**	-0.025	0.613	RFG	0.050	**4.110**	0.360	-0.045	4.261
	IFG	0.050	10.205	0.296	-0.030	0.644	IFG	0.050	4.147	0.356	-0.039	3.727
GRU	$r(e_t, z_j)$	0.025	**9.076**	**0.274**	**0.002**	-3.063	$r(e_t, z_j)$	0.025	**3.517**	**0.325**	**-0.003**	**0.923**
	EM	0.025	14.421	0.471	-0.157	123.946	EM	0.025	4.329	0.356	-0.111	9.201
	RFG	0.025	9.750	0.299	-0.009	**-0.337**	RFG	0.025	3.545	0.336	-0.051	3.926
	IFG	0.025	9.348	0.277	-0.031	2.499	IFG	0.025	3.564	0.333	-0.060	4.744
T-GCN	$r(e_t, z_j)$	0.025	**9.968**	**0.310**	**0.006**	**-5.551**	$r(e_t, z_j)$	0.200	**3.767**	0.410	**-0.001**	**0.898**
	EM	0.025	15.220	0.498	-0.188	153.571	EM	0.200	4.050	0.431	-0.048	7.186
	RFG	0.025	10.866	0.352	-0.049	13.691	RFG	0.200	3.990	0.394	-0.018	3.570
	IFG	0.025	10.857	0.338	-0.082	22.953	IFG	0.200	3.982	**0.370**	0.005	2.119
STGCN	$r(e_t, z_j)$	0.050	**8.455**	**0.236**	0.006	**-1.644**	$r(e_t, z_j)$	0.300	**4.062**	0.401	0.002	**0.835**
	EM	0.050	12.723	0.407	**-0.004**	21.668	EM	0.300	4.576	0.408	-0.059	6.618
	RFG	0.050	10.276	0.300	0.019	4.497	RFG	0.300	4.092	**0.362**	**0.000**	-2.710
	IFG	0.050	8.824	0.259	0.025	-11.961	IFG	0.300	4.403	0.382	-0.018	3.307
ConvLSTM	$r(e_t, z_j)$	0.025	8.314	**0.249**	**0.003**	**-0.736**	$r(e_t, z_j)$	0.025	**3.325**	0.349	**-0.001**	**0.497**
	EM	0.025	12.874	0.431	-0.161	102.158	EM	0.025	3.474	0.346	-0.097	6.164
	RFG	0.025	8.806	0.276	0.014	-8.626	RFG	0.025	3.437	0.325	-0.146	9.645
	IFG	0.025	**8.231**	0.254	-0.066	9.405	IFG	0.025	3.382	**0.316**	-0.069	4.705

flexibly adopt any model they want in addressing fairness issues. Also, the proposed method enjoys great transparency since end-users (e.g., stakeholders) can easily understand how fairness is being taken into account and improved (from the fairness regularization term). Moreover, this method is transferable for other forecasting applications. Besides travel demand forecasting, other important issues including traffic count forecasting, pedestrian activity forecasting or crash frequency forecasting may also have silent fairness problems. Researchers can apply our proposed method to address the fairness issues and provide fair decision-making. This study only examined the proposed method using time-series (panel) data. However, we believe it can be easily generalized to other applications with different data modalities. For example, transportation-planning models, which usually use cross-sectional data, should also be examined with fairness analysis. Our unfairness correction method can be flexibly adopted by planning models (e.g., [59]) to inform fair design of transportation ecosystems. Flexibility is also reflected in that, once the models are trained, access to protected attributes is no longer required. Unlike the post-processing technique that always requires access to the protected attribute [31], [36], our approach lifts this restriction and can be flexibly adapted for future forecasting tasks.

Third, *effectiveness in achieving fairness while preserving prediction accuracy*. Multiple studies reported that machine learning has a trade-off between accuracy and fairness (e.g., [28], [36]), i.e., the reduction of unfairness will inevitably trigger an accuracy drop. Our scheme addresses this trade-off by incorporating an interactive weight coefficient (i.e., λ) into the loss function. We treat λ as a hyperparameter of the learning tasks (i.e., improving fairness while preserving accuracy) and tune it together with other hyperparameters. In this way, the model automatically finds the optimal hyperparameter combination that has the best performance in improving fairness while maintaining prediction accuracy. Most of our experiments revealed that this approach could significantly reduce unfairness only at little expense of accuracy decline. While in some cases, our proposed method can even significantly improve fairness and slightly improve prediction accuracy.

B. Policy Implications

Dynamically balancing the supply and demand for transportation systems is important to improve cost-benefit effects and efficiency. And this balance relies heavily on accurate predictions [2]. Although machine learning intensively promotes predictions, it may simultaneously introduce bias. The overall satisfactory predictions may hide a huge prediction accuracy gap across areas of the city or underrepresented groups of residents [12], [13]. Our study also confirms this finding. Specifically, Table III shows that both machine learning and statistical models can produce lower prediction accuracy for the disadvantaged communities (i.e., the non-white-majority, the lower-education-attainment, the elderly and the low-income) than that of the advantaged communities. The predictive disparity implies that if transportation planners naively use such travel demand forecasting models without accounting for the fairness issues, the modeling results will lead to ineffective transportation resource allocations, impede the mobility of the disadvantaged communities, and even possibly further exacerbate the existing operational biases of ridesourcing services, e.g., higher trip-cancellation rate, longer waiting times and higher per-mile fees for disadvantaged communities [16], [17], [79], [80].

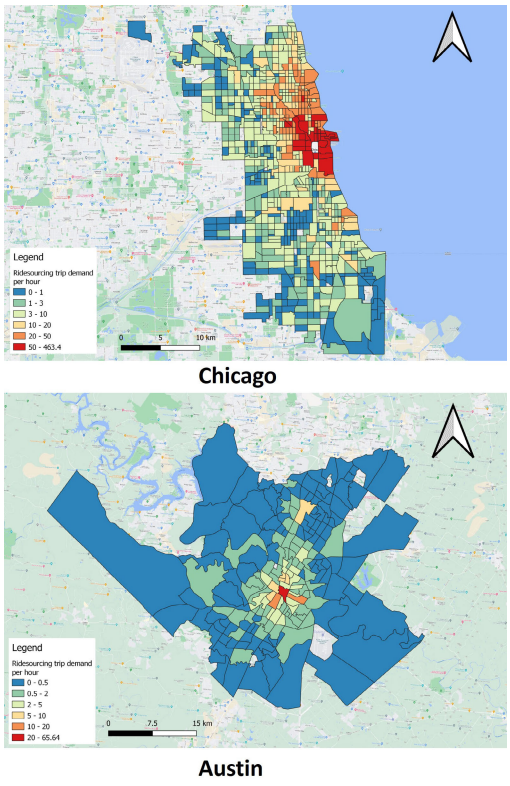

Fig. 1. Spatial distribution of the average ridesourcing demand per hour.

Our proposed method can help mitigate the unfairness issues of the current ridesourcing operations to better serve the disadvantaged communities. We believe that ridesourcing policymakers should consider incorporating our proposed method into the travel demand modeling framework to inform fairer ridesourcing resource allocations and operations. Additionally, two fairness metrics can be used by city governments to evaluate and regulate ridesourcing operations. Moreover, the fairness measurements and unfairness correction method can be adopted to facilitate the effective operations of other travel modes such as public transit and shared micromobility. For example, an accurate and fair demand forecasting model will enable transit authorities to provide more personalized transit services to balance operation efficiency and effectiveness [81]. Also, a fairness-aware travel demand forecasting model will help micromobility (e.g., bikeshare and e-scooter) operators better rebalance the vehicles and ensure fair distribution of service availability throughout the day [82]. As travel demand is a crucial element in cost-benefit analysis, our proposed fairness-aware travel demand forecasting model can also act as a guiding tool for improving evaluations of the economic sustainability of the mobility systems in projects [83], [84].

C. Limitations and Future Research Directions

This study has some limitations that warrant follow-up investigations. For example, we only evaluated the proposed methodology using two fairness metrics (i.e., prediction accuracy gap and correlation coefficient) in this paper. Future works may consider using a wider range of fairness metrics to conduct a comprehensive evaluation. Moreover, by using correlation techniques, we assume the prediction accuracy is linearly correlated with the protected attributes. Future studies may consider exploring whether this association is nonlinear and developing corresponding methods. Another widely debated research topic is the connection between accuracy and fairness. Several previous studies have shown that the accuracy-fairness trade-off exists across datasets and applications [28], [58] while others have shown that improvements in accuracy and fairness can co-occur [20]. Hence, forthcoming investigations may shed further light on this relationship, such as identifying scenarios in which fairness and accuracy can both be enhanced or where the accuracy-fairness trade-off is prominent. In addition, this study only examined one travel mode (i.e., ridesourcing). A more comprehensive analysis that includes various travel modes (e.g., transit, car-sharing, and shared micromobility) and diverse contexts (e.g., different locations) should be conducted to test the generalizability and robustness of the unfairness correction method. Also, we only considered the first step (i.e., trip generation) of the four-step travel demand models. A more comprehensive travel demand model should also include trip distribution (origin-destination demand) estimation [51], [59], [85]. Therefore, future studies may consider addressing fairness issues in the distribution of travel demand as well. Finally, we note that this study is for empirical prediction rather than a causal analysis [68], [86], i.e., we are not investigating which factors are causing and how much they contribute to the unfairness. However, we acknowledge that such causal analysis should be an important component for a complete picture of fair travel demand modeling, and thus needs to be considered by future studies.

VII. Conclusion

This study examines the fairness issues in travel demand forecasting models and develops a new methodology to enhance their fairness while preserving the prediction accuracy. By leveraging two real-world ridesourcing-trip data from Chicago, IL and Austin, TX, the unfairness issues of seven state-of-the-art AI-based models on forecasting travel demand are evaluated. A novel and transparent in-processing method, which is based on an absolute correlation regularization term, is proposed to simultaneously address the unfairness arising from multiple protected attributes. We also compare the performance (including both fairness and accuracy) of our proposed unfairness correction method with three state-of-the-art unfairness correction methods to show its effectiveness.

The results highlight that both statistical and machine learning models have pronounced fairness issues, wherein the prediction accuracy for advantaged groups are notably higher than disadvantaged groups. Our proposed unfairness correction method can effectively enhance fairness for multiple protected attributes while preserving prediction accuracy. The comparative study reveals that our proposed method significantly outperforms other methods in both fairness and accuracy. Beyond its performance, our proposed method has remarkable flexibility—it is model-agnostic and can be adapted to various applications and different data modality. In summary, this study advances our understanding of fairness issues in travel

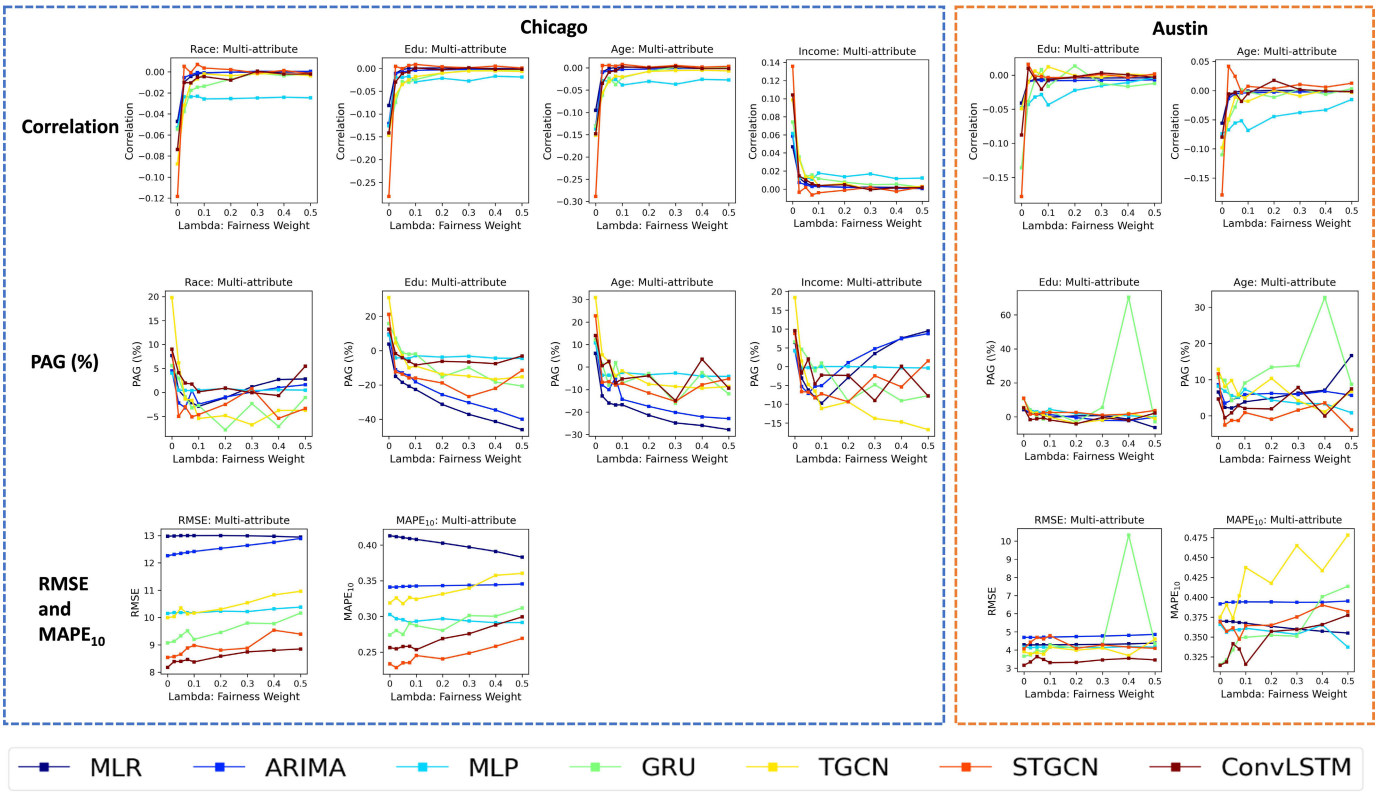

Fig. 2. Sensitivity analysis of λ across two case studies.

TABLE IX
CORRELATION MATRIX OF PROTECTED ATTRIBUTES (CHICAGO)

	Race	Edu	Age	Income
Race	1.000	0.629	0.490	-0.746
Edu	0.629	1.000	0.679	-0.620
Age	0.490	0.679	1.000	-0.407
Income	-0.746	-0.620	-0.407	1.000

TABLE X
CORRELATION MATRIX OF PROTECTED ATTRIBUTES (AUSTIN)

	Race	Edu	Age	Income
Race	1.000	0.605	-0.178	-0.401
Edu	0.605	1.000	-0.062	-0.424
Age	-0.178	-0.062	1.000	0.620
Income	-0.401	-0.424	0.620	1.000

demand forecasting and equips transportation researchers with a powerful tool to foster fairness within the transportation ecosystem.

APPENDIX A

Fig. 1 shows the average ridesourcing demand per hour in Chicago, IL and Austin, TX. The spatial unit is the census tract. This plot reveals an evident disparity regarding ridesourcing demand across different areas.

APPENDIX B

The pairwise correlation matrix of the selected four protected attributes for two case studies is shown in Table. IX and Table X. Results show that the protected attributes are evidently correlated with each other.

APPENDIX C

The results of the sensitivity analysis for fairness weight, i.e., λ are presented in Fig. 2. We specifically investigated the effects of λ on model's prediction accuracy by RMSE and MAPE_{10} and fairness by both PAG and correlations. Note: In Austin case, setting fairness weight as 0.4 for GRU led to a substantial increase in RMSE and PAG. One possible reason could be that this combination of hyperparameters might explode the gradients and thus lead to this numerical instability.

ACKNOWLEDGMENT

During the preparation of this work, the authors used ChatGPT to check grammar errors and improve languages. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

[1] H. Xu, T. Zou, M. Liu, Y. Qiao, J. Wang, and X. Li, "Adaptive spatiotemporal dependence learning for multi-mode transportation demand prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 18632–18642, Oct. 2022.

[2] K.-F. Chu, A. Y. S. Lam, and V. O. K. Li, "Deep multi-scale convolutional LSTM network for travel demand and origin-destination predictions," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 8, pp. 3219–3232, Aug. 2020.

[3] Y. Xu, X. Zhao, X. Zhang, and M. Paliwal, "Real-time forecasting of dockless scooter-sharing demand: A spatio-temporal multi-graph transformer approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 8, pp. 8507–8518, Aug. 2023, doi: 10.1109/TITS.2023.3239309.

[4] H. Chen and Y. Cheng, "Travel mode choice prediction using imbalanced machine learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 4, pp. 3795–3808, Apr. 2023.

[5] A. Beutel et al., "Putting fairness principles into practice: Challenges, metrics, and improvements," in *Proc. AAAI/ACM Conf. AI, Ethics, Soc.*, Jan. 2019, pp. 453–459.

[6] R. S. Baker and A. Hawn, "Algorithmic bias in education," *Int. J. Artif. Intell. Educ.*, vol. 32, no. 4, pp. 1052–1092, Dec. 2022.

[7] J. Angwin, J. Larson, S. Mattu, and L. Kirchner, "Machine bias," in *Ethics of Data and Analytics*. New York, NY, USA: Auerbach Publications, 2016, pp. 254–264.

[8] C. Barabas, M. Virza, K. Dinakar, J. Ito, and J. Zittrain, "Interventions over predictions: Reframing the ethical debate for actuarial risk assessment," in *Proc. 1st Conf. Fairness, Accountability Transparency*, 2018, pp. 62–76.

[9] M. O. R. Prates, P. H. Avelar, and L. C. Lamb, "Assessing gender bias in machine translation: A case study with Google translate," *Neural Comput. Appl.*, vol. 32, no. 10, pp. 6363–6381, May 2020.

[10] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," *Science*, vol. 366, no. 6464, pp. 447–453, Oct. 2019.

[11] Z. Wei, Z. Li, M. M. Kulkarni, and X. Yue, "Equitable traffic crash prediction framework to support safety improvement grants allocation," *Inst. Transp. Engineers*, vol. 93, no. 9, pp. 37–45, 2023.

[12] A. Yan and B. Howe, "Fairness-aware demand prediction for new mobility," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, 2020, pp. 1079–1087.

[13] Y. Zheng, S. Wang, and J. Zhao, "Equality of opportunity in travel behavior prediction with deep neural networks and discrete choice models," *Transp. Res. C, Emerg. Technol.*, vol. 132, Nov. 2021, Art. no. 103410.

[14] X. Guo, H. Xu, D. Zhuang, Y. Zheng, and J. Zhao, "Fairness-enhancing vehicle rebalancing in the ride-hailing system," 2023, *arXiv:2401.00093*.

[15] A. Pandey and A. Caliskan, "Disparate impact of artificial intelligence bias in ridehailing economy's price discrimination algorithms," in *Proc. AAAI/ACM Conf. AI, Ethics, Soc.*, Jul. 2021, pp. 822–833.

[16] A. Brown, "Not all fees are created equal: Equity implications of ride-hail fee structures and revenues," *Transp. Policy*, vol. 125, pp. 1–10, Sep. 2022.

[17] H. Yang, Y. Liang, and L. Yang, "Equitable? Exploring ridesourcing waiting time and its determinants," *Transp. Res. D, Transp. Environ.*, vol. 93, Apr. 2021, Art. no. 102774.

[18] Y. Berhanu, E. Alemayehu, and D. Schröder, "Examining car accident prediction techniques and road traffic congestion: A comparative analysis of road safety and prevention of world challenges in low-income and high-income countries," *J. Adv. Transp.*, vol. 2023, pp. 1–18, Jul. 2023.

[19] M. Vega-Gonzalo and P. Christidis, "Fair models for impartial policies: Controlling algorithmic bias in transport behavioural modelling," *Sustainability*, vol. 14, no. 14, p. 8416, Jul. 2022.

[20] A. Yan, *Fairness-Aware Spatio-Temporal Prediction for Cities*. Seattle, WA, USA: Univ. Washington Press, 2021.

[21] S. He and K. G. Shin, "Socially-equitable interactive graph information fusion-based prediction for urban dockless E-scooter sharing," in *Proc. ACM Web Conf.*, Apr. 2022, pp. 3269–3279.

[22] D. Kaur, S. Uslu, K. J. Rittichier, and A. Durresi, "Trustworthy artificial intelligence: A review," *ACM Comput. Surv.*, vol. 55, no. 2, pp. 1–38, 2022.

[23] B. Li et al., "Trustworthy AI: From principles to practices," *ACM Comput. Surv.*, vol. 55, no. 9, pp. 1–46, Jan. 2023.

[24] M. Wan, D. Zha, N. Liu, and N. Zou, "In-processing modeling techniques for machine learning fairness: A survey," *ACM Trans. Knowl. Discovery From Data*, vol. 17, no. 3, pp. 1–27, Apr. 2023.

[25] A. Bose and W. Hamilton, "Compositional fairness constraints for graph embeddings," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2019, pp. 715–724.

[26] J. Dressel and H. Farid, "The accuracy, fairness, and limits of predicting recidivism," *Sci. Adv.*, vol. 4, no. 1, Jan. 2018, Art. no. eaao5580.

[27] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM Comput. Surv.*, vol. 54, no. 6, pp. 1–35, Jul. 2021.

[28] R. Berk et al., "A convex framework for fair regression," 2017, *arXiv:1706.02409*.

[29] S. Barocas and A. D. Selbst, "Big data's disparate impact," *California Law Rev.*, vol. 104, p. 671, Sep. 2016.

[30] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel, "Fairness through awareness," in *Proc. 3rd Innov. Theor. Comput. Sci. Conf.*, Jan. 2012, pp. 214–226.

[31] M. Hardt, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, Dec. 2016, pp. 3323–3331.

[32] T. Calders, A. Karim, F. Kamiran, W. Ali, and X. Zhang, "Controlling attribute effect in linear regression," in *Proc. IEEE 13th Int. Conf. Data Mining*, Dec. 2013, pp. 71–80.

[33] M. Kusner, J. Loftus, C. Russell, and R. Silva, "Counterfactual fairness," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, Dec. 2017, pp. 4069–4079.

[34] F. Kamiran and T. Calders, "Data preprocessing techniques for classification without discrimination," *Knowl. Inf. Syst.*, vol. 33, no. 1, pp. 1–33, Oct. 2012.

[35] F. P. Calmon, D. Wei, B. Vinzamuri, K. N. Ramamurthy, and K. R. Varshney, "Optimized pre-processing for discrimination prevention," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, Dec. 2017, pp. 3995–4004.

[36] A. Agarwal, M. Dudík, and Z. S. Wu, "Fair regression: Quantitative definitions and reduction-based algorithms," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 120–129.

[37] K. D. Johnson, D. P. Foster, and R. A. Stine, "Impartial predictive modeling: Ensuring fairness in arbitrary models," 2016, *arXiv:1608.00528v2*.

[38] S. Caton and C. Haas, "Fairness in machine learning: A survey," 2020, *arXiv:2010.04053*.

[39] A. Wang and O. Russakovsky, "Directional bias amplification," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 10882–10893.

[40] J. Yang, A. A. S. Soltan, D. W. Eyre, Y. Yang, and D. A. Clifton, "An adversarial training framework for mitigating algorithmic biases in clinical machine learning," *NPJ Digit. Med.*, vol. 6, no. 1, p. 55, Mar. 2023.

[41] D. Xu, Y. Wu, S. Yuan, L. Zhang, and X. Wu, "Achieving causal fairness through generative adversarial networks," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 1452–1458.

[42] A. Yan and B. Howe, "Equitensors: Learning fair integrations of heterogeneous urban data," in *Proc. Int. Conf. Manage. Data*, 2021, pp. 2338–2347.

[43] P. Cheng, W. Hao, S. Yuan, S. Si, and L. Carin, "FairFil: Contrastive neural debiasing method for pretrained text encoders," 2021, *arXiv:2103.06413*.

[44] N. Quadrianto, V. Sharmanska, and O. Thomas, "Discovering fair representations in the data domain," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8219–8228.

[45] M. Du, F. Yang, N. Zou, and X. Hu, "Fairness in deep learning: A computational perspective," *IEEE Intell. Syst.*, vol. 36, no. 4, pp. 25–34, Aug. 2020.

[46] B. H. Zhang, B. Lemoine, and M. Mitchell, "Mitigating unwanted biases with adversarial learning," in *Proc. AAAI/ACM Conf. AI Ethics Soc.*, 2018, pp. 335–340.

[47] T. Kamishima, S. Akaho, and J. Sakuma, "Fairness-aware learning through regularization approach," in *Proc. IEEE 11th Int. Conf. Data Mining Workshops*, Dec. 2011, pp. 643–650.

[48] H. Kim et al., "Counterfactual fairness with disentangled causal effect variational autoencoder," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 9, 2021, pp. 8128–8136.

[49] M. Kearns, S. Neel, A. Roth, and Z. S. Wu, "Preventing fairness gerrymandering: Auditing and learning for subgroup fairness," in *Proc. 35th Int. Conf. Mach. Learn.*, Jul. 2018, pp. 2564–2572.

[50] M. Kearns, S. Neel, A. Roth, and Z. S. Wu, "An empirical study of rich subgroup fairness for machine learning," in *Proc. Conf. Fairness, Accountability, Transparency*, Jan. 2019, pp. 100–109.

[51] D. T. Hartgen, "Hubris or humility? Accuracy issues for the next 50 years of travel demand modeling," *Transportation*, vol. 40, no. 6, pp. 1133–1157, Nov. 2013.

[52] G. Guo and T. Zhang, "A residual spatio-temporal architecture for travel demand forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 115, Jun. 2020, Art. no. 102639.

[53] D. Michailidis, W. Röpke, S. Ghebreab, D. M. Roijers, and F. P. Santos, "Fairness in transport network design—A multi-objective reinforcement learning approach," in *Proc. Adapt. Learn. Agents Workshop AAMAS*, 2023, pp. 1–7.

[54] K. Kerkman, K. Martens, and H. Meurs, "Predicting travel flows with spatially explicit aggregate models: On the benefits of including spatial dependence in travel demand modeling," *Transp. Res. A, Policy Pract.*, vol. 118, pp. 68–88, Dec. 2018.

[55] D. Jalota, K. Solovey, M. Tsao, S. Zoepf, and M. Pavone, "Balancing fairness and efficiency in traffic routing via interpolated traffic assignment," *Auto. Agents Multi-Agent Syst.*, vol. 37, no. 2, p. 32, Dec. 2023.

[56] O. Jahn, R. H. Möhring, A. S. Schulz, and N. E. Stier-Moses, "System-optimal routing of traffic flows with user constraints in networks with congestion," *Oper. Res.*, vol. 53, no. 4, pp. 600–616, Aug. 2005.

[57] D. Duran-Rodas, B. Wright, F. C. Pereira, and G. Wulfhorst, "Demand And/oR equity (DARE) method for planning bike-sharing," *Transp. Res. D, Transp. Environ.*, vol. 97, Aug. 2021, Art. no. 102914.

[58] A. Chouldechova and A. Roth, "The frontiers of fairness in machine learning," 2018, *arXiv:1810.08810*.

[59] X. Zhang and X. Zhao, "Machine learning approach for spatial modeling of ridesourcing demand," *J. Transp. Geography*, vol. 100, Apr. 2022, Art. no. 103310.

[60] A. Olteanu, C. Castillo, F. Diaz, and E. Kiciman, "Social data: Biases, methodological pitfalls, and ethical boundaries," *Frontiers Big Data*, vol. 2, p. 13, Jul. 2019.

[61] A. E. Brown, "Prevalence and mechanisms of discrimination: Evidence from the ride-hail and taxi industries," *J. Planning Educ. Res.*, vol. 43, no. 2, pp. 268–280, Jun. 2023.

[62] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecasting*, vol. 22, no. 4, pp. 679–688, Oct. 2006.

[63] C. Chen, J. Twycross, and J. M. Garibaldi, "A new accuracy measure based on bounded relative error for time series forecasting," *PLoS One*, vol. 12, no. 3, Mar. 2017, Art. no. e0174202.

[64] F. S. Tabataba et al., "A framework for evaluating epidemic forecasts," *BMC Infectious Diseases*, vol. 17, no. 1, pp. 1–27, Dec. 2017.

[65] L. Tornqvist, P. Vartia, and Y. O. Vartia, "How should relative changes be measured?" *Amer. Statistician*, vol. 39, no. 1, p. 43, Feb. 1985.

[66] S. Corbett-Davies, E. Pierson, A. Feller, S. Goel, and A. Huq, "Algorithmic decision making and the cost of fairness," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2017, pp. 797–806.

[67] Z. Bai and P. Krishnaiah, "Reduction of dimensionality," in *Encyclopedia of Physical Science and Technology*, 3rd ed., R. A. Meyers, Ed. New York, NY, USA: Academic, 2003, pp. 55–73. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B012227410500466X

[68] G. Shmueli, "To explain or to predict?" *Stat. Sci.*, vol. 25, no. 3, pp. 289–310, 2010.

[69] P. D. Allison, *Multiple Regression: A Primer*. Newbury Park, CA, USA: Pine Forge Press, 1999.

[70] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Struct. Multidiscip. Optim.*, vol. 26, no. 6, pp. 369–395, 2004.

[71] O. Sener and V. Koltun, "Multi-task learning as multi-objective optimization," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, Dec. 2018, pp. 525–536.

[72] S. Makridakis and M. Hibon, "ARMA models and the Box–Jenkins methodology," *J. Forecast.*, vol. 16, no. 3, pp. 147–163, 1997

[73] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," 2014, *arXiv:1409.1259*.

[74] L. Zhao et al., "T-GCN: A temporal graph convolutional network for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3848–3858, Aug. 2019.

[75] X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst.*, vol. 1, Dec. 2015, pp. 802–810.

[76] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," 2017, *arXiv:1709.04875*.

[77] J. Tang, J. Liang, F. Liu, J. Hao, and Y. Wang, "Multi-community passenger demand prediction at region level based on spatio-temporal graph convolutional network," *Transp. Res. C, Emerg. Technol.*, vol. 124, Mar. 2021, Art. no. 102951.

[78] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[79] A. Brown et al., "The equalizer: Could ride-hailing extend equitable car access?" *Transfers Mag.*, vol. 4, p. 2, Jul. 2019.

[80] A. Brown and R. Williams, "Equity implications of ride-hail travel during COVID-19 in California," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2677, no. 4, pp. 1–14, Apr. 2023.

[81] A. Ermagun and N. Tilahun, "Equity of transit accessibility across Chicago," *Transp. Res. D, Transp. Environ.*, vol. 86, Sep. 2020, Art. no. 102461.

[82] X. Yan, W. Yang, X. Zhang, Y. Xu, I. Bejleri, and X. Zhao, "A spatiotemporal analysis of e-scooters' relationships with transit and station-based bikeshare," *Transp. Res. D, Transp. Environ.*, vol. 101, Dec. 2021, Art. no. 103088.

[83] M. De Aloe, R. Ventura, M. Bonera, B. Barabino, and G. Maternini, "Applying cost-benefit analysis to the economic evaluation of a tram-train system: Evidence from Brescia (Italy)," *Res. Transp. Bus. Manage.*, vol. 47, Mar. 2023, Art. no. 100916.

[84] R. Ventura, M. Bonera, M. Carra, B. Barabino, and G. Maternini, "Evaluating the viability of a tram-train system. A case study from Salento (Italy)," *Case Stud. Transp. Policy*, vol. 10, no. 3, pp. 1945–1963, Sep. 2022.

[85] X. Zhang, Z. Zhou, Y. Xu, and X. Zhao, "Analyzing spatial heterogeneity of ridesourcing usage determinants using explainable machine learning," *J. Transp. Geography*, vol. 114, Jan. 2024, Art. no. 103782.

[86] H. E. Brady, "Causation and explanation in social science," in *The Oxford Handbook of Political Science*. London, U.K.: Oxford Univ. Press, 2011, doi: 10.1093/oxfordhb/9780199604456.013.0049.

Xiaojian Zhang received the B.E. degree from Southwest Jiaotong University, China. He is currently pursuing the Ph.D. degree with the Department of Civil and Coastal Engineering, University of Florida, Gainesville, USA. He is working on promoting fairness in travel behavior analysis and developing new AI methods to tackle challenging problems at the intersection of transportation and emergency management.

Qian Ke received the B.E. degree in statistics from Huazhong University of Science and Technology, Wuhan, China, in 2014, and the first M.S.E. degree in applied mathematics and statistics, the second M.S.E. degree in computer science, and the Ph.D. degree in statistics from Johns Hopkins University, Baltimore, MD, USA, in 2015, 2019, and 2020, respectively. She is currently an AI Researcher with Bloomberg. Her work focuses on developing and applying machine learning methods to revolutionize time series forecasting within the financial sector.

Xilei Zhao received the B.E. degree in civil engineering from Southeast University, Nanjing, China, in 2013, and the M.S.E. degree in civil engineering and in applied mathematics and statistics and the Ph.D. degree in civil engineering from Johns Hopkins University, Baltimore, MD, USA, in 2016 and 2017, respectively. She is currently an Assistant Professor with the Department of Civil and Coastal Engineering, University of Florida, where she leads the Smart, Equitable, Resilient Mobility Systems (SERMOS) Laboratory. Prior to her appointment with the University of Florida, she was a Postdoctoral Fellow with the School of Industrial and Systems Engineering, Georgia Tech, from 2018 to 2019; and a Research Fellow with the Department of Industrial and Operations Engineering, University of Michigan, from 2017 to 2018. Her work focuses on developing and applying data and computational science methods to tackle challenging problems in transportation and resilience. She specializes in big data analytics and trustworthy AI applications in travel behavior modeling; modeling and planning evacuation; and quantifying resilience for critical infrastructure systems, societal systems, and communities.