UNDERSTANDING TRAVEL BEHAVIOR: A DEEP NEURAL NETWORK AND SHAP APPROACH TO MODE CHOICE DETERMINANTS

H. Çevik,* *O. Přibyl*,* *S. Samandar*†

Abstract: Understanding individual travel behavior is crucial for developing effective travel demand management strategies and informed transportation policies. This study investigates the factors influencing individuals' mode choices by analyzing data from a comprehensive travel survey. We employ a deep neural network model to explore the relationships between survey variables and respondents' transportation mode preferences, focusing on both observable and latent factors. The SHAP method is applied to interpret the model's outputs, providing global and local explanations that offer detailed insights into the contribution of each variable to mode choice decisions. By identifying the key determinants of mode selection and uncovering the complex interactions between these factors, this research provides valuable insights for designing targeted policies that can better address transportation needs and influence sustainable travel behavior.

Key words: *mode choice, travel behavior, interpretable AI, XAI, SHAP, DNN*

Received: June 5, 2024 **DOI:** 10.14311/NNW.2024.34.012
Revised and accepted: September 2, 2024

1. Introduction

Mode choice analysis plays a crucial role in transportation planning and policy development, as it helps to understand and forecast travel demands [2]. Travelers adopt new decisions and rely on environmental factors as key elements in their repeated decision-making process, particularly in stable contexts [26]. Furthermore, socioeconomic factors such as gender, age, educational level, and income level significantly influence mode choice decisions [20]. This research, therefore, examines the important factors underlying how travelers make decisions regarding their mode choice.

While statistical regression and discrete choice models have been commonly used for travel mode analysis, they may offer some advantages in terms of results

*Halil Çevik – Corresponding author; Ondřej Přibyl; Czech Technical University in Prague, Faculty of Transportation Sciences, Na Florenci 25, Praha 1, 110 00, Czech Republic, E-mail: cevikhal@cvut.cz

†Shoaib Samandar; North Carolina State University, Institute for Transportation Research and Education, Raleigh, North Carolina, 27606-3870, U.S.A.

interpretability but have limitations in capturing unobserved factors and complex relationships [9]. Nonlinear modeling approaches, such as support vector machines and decision trees, can provide more effective mode choice analysis compared to traditional statistical methods [22]. Advances in deep learning have produced complex, high-performing models that lack interpretability, making it difficult for users to trust the models and use the feature importance to enhance performance [9]. When using deep neural networks for mode choice modeling, black-box explanations might be insufficient. Therefore, we employed an interpretable AI method to understand why and how individuals make their travel mode choices.

Clarification of the terms "interpretation" and "explanation" is necessary since they will be used throughout the paper and serve the purpose of this work. As defined in [21], an interpretation is the mapping of an abstract concept into a form that humans can understand, and an explanation is the set of interpretable features that contributed to producing a particular decision, such as mode choice decision, for a given example. We focus on interpreting the results of a deep neural network and explaining each input variable as well as the relationships among them. This work employed a methodology to analyze mode choices and provide a clearer understanding of the results by using interpretable AI approaches within deep neural network architectures.

This paper addresses two key problems. The first is looking at travel behavior and understanding the nonlinear and complex relationships between variables, as well as exploring the factors that influence it. The second is the need for deep neural networks due to the simplified structure of statistical methods, with feature ranking alone being insufficient. However, DNN models introduce a black-box nature that requires explanation. Thus, the interpretability of these complex model structures is another challenge that must be addressed.

In order to explore the decision-making process regarding mode choices, a travel survey served as the input for our deep neural network model. The primary goal is to provide interpretations of the DNN model's results, revealing the reasons behind respondents' selection of specific travel modes. The target variable of interest is the mode choice, and the model interpretation will shed light on how each feature represents and influences the outcome, as well as the relevance of these features in explaining the results. This paper seeks to address these issues through the activity diagram shown in Fig. 2.

The paper is organized as follows: Section 2 presents the state-of-the-art mode choice analysis and interpretability methods for understanding model behavior. Section 3 describes the methodological approach, including the deep neural network model used for mode choice analysis, the SHAP methodology applied to interpret the model results, and the stability of SHAP explanations. Section 4 presents the results from a case study detailing the modeling framework, the dataset, and the interpretations of the model outputs. Finally, Section 5 discusses the findings in the context of existing research, and Section 6 concludes the paper by highlighting key contributions and future research directions.

2. State of the Art

This section outlines an overview of machine learning models used for mode choice analysis, introduces the crucial concept of model interpretability, and presents various techniques to interpret both machine learning and deep learning models.

2.1 Methods for Mode Choice Analysis

In urban transportation planning, knowing the transport mode is a crucial aspect, which is typically explored through questionnaires/travel diaries/ telephone interviews [8]. Transport mode choice modeling methods include K-nearest neighbor, support vector machines, and tree-based approaches such as single decision tree, bagging, and random forest [8]. Additionally, deep learning models have proven to be effective tools for analyzing transport mode choice, as they have outperformed traditional discrete choice models in terms of predictive accuracy for both individual and aggregated behavior [24].

Discrete choice models like the multinomial logit have long been used to analyze individual travel behavior among discrete alternatives, despite their simplistic assumption of linear utility for complex human choices [33]. However, they remain practical due to their interpretability [33]. Meanwhile, data-driven machine learning techniques, particularly deep learning models, are emerging as alternatives in transportation research, in contrast to the traditional multinomial logit model [33].

In travel behavior research, the main points discussed when comparing machine learning and discrete choice modeling methods are their predictive performance and their capacity to offer insights into travelers' decision-making processes [18]. While prediction is a common objective in all modeling approaches, whether using discrete choice models or machine learning classifiers, many transportation applications also require the ability to interpret the findings [31]. Achieving high prediction accuracy often requires using complex non-linear models like support vector machines, deep neural networks, or random forests [7]. However, these sophisticated models tend to lack interpretability [7].

A key concern with using deep neural networks for transportation mode choice modeling is their perceived lack of interpretability [31]. Interpretation tools have been proposed to explain or reveal how deep models make decisions, as deep neural networks are often seen as less interpretable due to their over-parameterized "black box" model structure [12]. Prior studies employing deep neural networks for transportation choice modeling have primarily focused on using DNN to predict mode choice, activity choice, car ownership, and other related choices [31]. Only a few transportation studies have examined the interpretability of DNN in choice modeling without providing explicit metrics to measure the quality of interpretability [31]. In comparison to discrete choice models, the interpretability of DNN models will be a key factor in determining whether these approaches can be used for demand prediction in transportation contexts and have practical implications on our understanding of individual decision-making behavior [31].

2.2 Interpretable AI

Interpretability is crucial for gaining insights into mode choice behavior and offering these insights to inform urban mobility patterns [10]. Understanding the model's interpretability is important for users to have better trust in the model and better comprehend the significance of the features for further improving the model's performance [9]. There are several techniques to interpret machine learning and deep learning models. These include Shapley additive explanation (SHAP), local interpretable model-agnostic explanation (LIME), deep learning important features (DeepLIFT), model agnostic concept extractor (MACE), and generative adversarial network (GAN) based methods [25].

In this work, we used Shapley additive explanations due to its model-agnostic nature, which allows it to be applied to explain and analyze results across a wide range of machine learning and deep learning models [25]. SHAP has been successfully employed with various models such as random forest [5, 9, 29, 30, 11], logistic regression [9, 29], decision tree [9, 11], Naive Bayes [29], LSTM [25], XGBoost [30, 13], deep neural network [1]. The flexibility and broad applicability of SHAP make it a well-suited tool for interpreting model outputs and generating valuable insights into travel behavior.

3. Methods

This section outlines the methodological approaches used in this paper, including a technique for enhancing the interpretability of the deep neural network models. It presents the SHAP methodology and the calculation of SHAP values, with a focus on the deep explainer approach. It briefly introduces the theoretical background and provides a discussion on the stability and reliability of the obtained outcomes.

3.1 SHAP Methodology

SHAP (Shapley additive explanation) values are proposed by [17] as a unified measure of feature importance. Shapley sampling values aim to explain any model by applying sampling approximations to Eq. (1) and approximating the effect of removing a variable from the model through integration over samples from the training dataset [17].

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_{S\cup\{i\}}(x_{S\cup\{i\}}) - f_S(x_S)], \tag{1}$$

where ϕ_i, also known as the SHAP value, is the unified measure of additive feature attributions. The set of all features is denoted as F, and S represents the feature subsets. Models $f_{S\cup\{i\}}$ and f_S are trained with and without a feature, respectively. Then, predictions from the two models are compared on the current input $f_{S\cup\{i\}}(x_{S\cup\{i\}}) - f_S(x_S)$, where x_S represents the values of the input features in the set S. Since the impact of excluding a feature depends on other features in the model, the differences are computed for all possible subsets $S \subseteq F \setminus \{i\}$ [17].

SHAP values show how each feature affects the model's output by measuring the change in the expected prediction when considering that feature. They explain the output of a function f as the sum of the effects ϕi of each feature [16, 17]. The x-axis (see Fig. 1) shows the cumulative model output as features are progressively added. Each ϕi represents the contribution of the corresponding feature, and the length of the arrows shows the magnitude of each feature's effect on the prediction. Larger arrows indicate that the feature contributes more to the change in the model's prediction, while smaller arrows indicate a lesser contribution.

Despite the high accuracy of machine/deep-learning models, their results can be challenging for users to understand [25]. Additionally, it can be problematic to uncover hidden biases within the datasets or identify model weaknesses without a clear understanding of the decision-making process [25].

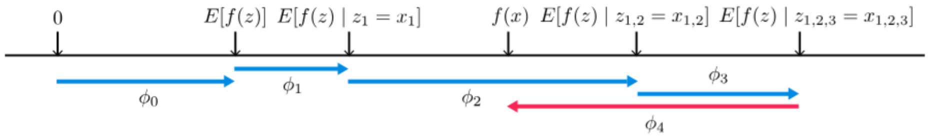

Fig. 1 *Contribution of features to model output. [17]*

3.2 Deep SHAP (DeepExplainer)

Deep SHAP combines deep learning important features (DeepLIFT) and Shapley values. DeepLIFT is introduced by [28], a novel algorithm that assigns importance scores to the input features for a given model output. DeepLIFT uses a backpropagation approach, which efficiently propagates an importance signal from an output neuron backward through the layers to the input in a single pass [28]. By assuming independence among the input features and linearity within the deep model, DeepLIFT approximates SHAP values [17].

Eqs. (2) through (5) below detail the calculations for SHAP values using the Deep SHAP method [17]:

$$m_{x_j f_3} = \frac{\phi_i(f_3, x)}{x_j - E[x_j]}, \tag{2}$$

$$\forall j \in \{1,2\}, \quad m_{y_i f_j} = \frac{\phi_i(f_j, y)}{y_i - E[y_i]}, \tag{3}$$

$$m_{y_i f_3} = \sum_{j=1}^{2} m_{y_i f_j} m_{x_j f_3} \quad \text{chain rule}, \tag{4}$$

$$\phi_i(f_3, y) \approx m_{y_i f_3}(y_i - E[y_i]) \quad \text{linear approximation}, \tag{5}$$

where

- m – multiplier that normalizes the SHAP value by the difference between the actual and expected values,
- ϕ – SHAP value, representing the contribution of a specific feature to the model's output,

- x – the actual value of the input feature,
- $E[x]$ – expected value of the input feature,
- y – actual output value of the model,
- $E[y]$ – expected output value.

Eq. (2) calculates the multiplier for feature j by normalizing the SHAP value of feature i with respect to the difference between the feature's value and its expected value. Similarly, Eq. (3) calculates the multiplier for the output (y_i). by normalizing the SHAP value of feature i with respect to the difference between the actual output and its expected value. Eq. (4) expresses the overall multiplier for the output (y_i) as the sum of the products of the multipliers from individual features j and their contributions to the overall model output (f_3). Finally, Eq. (5) provides an approximation of the SHAP value for feature i based on the model output and the difference between the actual output and its expected value, using the multiplier derived from previous calculations.

3.3 Stability of SHAP Explanations

The stability of SHAP explanations is positively correlated with the size of the background sample, which refers to a set of representative data points from the training dataset [32]. A larger sample reduces randomness, resulting in a more representative, stable, and reliable dataset [32]. Users are encouraged to use as large a background dataset as possible, potentially even an entire training dataset. However, it is essential to recognize that larger background datasets can significantly increase computational costs [32]. The optimal background dataset size ultimately depends on the desired level of accuracy in ranking variable importance [32].

Once the background dataset size is set, it must accurately represent the entire dataset, as this provides the SHAP explainer with key information about the population, which directly impacts the SHAP values [14]. When the majority class dominates, an unbalanced dataset sets a lower baseline for minority class observations, which can lead to an overestimation of their SHAP values, as SHAP measures the difference between the prediction and the average value [14].

While SHAP is a reliable approach for evaluating the importance of variables, the precise ranking of variables requires thoughtful consideration [32]. Furthermore, SHAP is more dependable in ranking the most and least significant variables compared to the moderately important ones [32].

4. Case Study

This section begins by outlining the modeling framework employed in the research, followed by a detailed description of the input data, including both explanatory and latent variables used in the model. Finally, the section presents the model results and their interpretation, using various SHAP visualizations such as summary plots, beeswarm plots, and decision plots to explain the influence of different variables on transport mode choice.

4.1 Description of The Modeling Framework

The framework, illustrated in Fig. 2, is represented as a unified modeling language (UML) activity diagram, outlining the step-by-step methodology. The process begins with data preprocessing, including cleaning the travel survey data and performing exploratory data analysis, which reveals general patterns in the data. The output variables indicating mode choices, originally categorical and labeled as public transport, mobility as a service (MaaS), and Car, were transformed using one-hot encoding. This process turned each choice into a unique binary format (e.g., 100, 010, and 001). This transformation was necessary to ensure that the deep learning model treats the mode choices as distinct categories rather than as ordinal or continuous values.

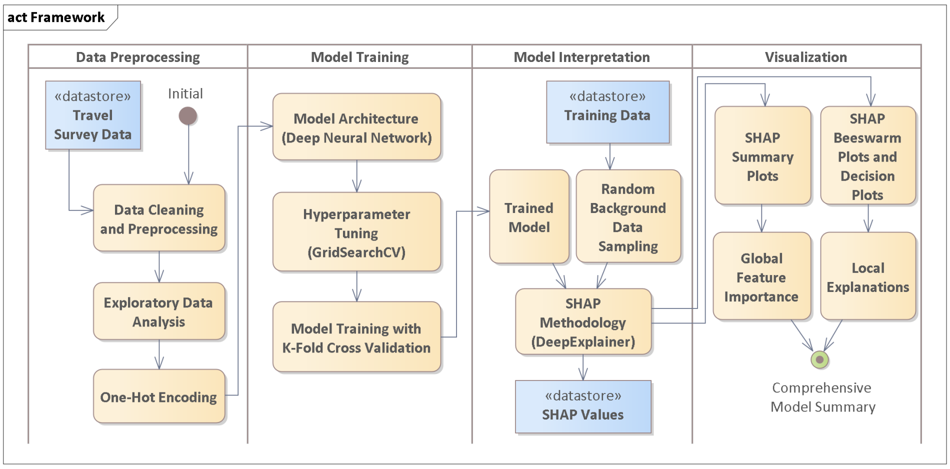

Fig. 2 *Research framework.*

During the model training phase, we developed a feedforward neural network model using the Keras library. The input layer consists of 21 neurons corresponding to the 21 input features. The network includes two hidden layers, each composed of 32 neurons. These hidden layers are fully connected, meaning each neuron is linked to every neuron in the next layer, allowing the model to capture complex patterns in the data. Finally, the output layer has 3 neurons, corresponding to the number of target classes. Given the dataset's moderate size and feature count, this network is designed to efficiently learn from the data and provide accurate interpretations of the model's outcome. To improve the model's reliability, we employed K-fold cross-validation during training, dividing the data into multiple subsets to ensure consistent performance across different data partitions.

After defining the model architecture, we used GridSearchCV to determine the optimal hyperparameters, such as the activation function, epoch number, batch size, optimizer, learning rate, and dropout ratio. This hyperparameter tuning process allowed us to obtain the optimal configuration and increase the model's performance. Tab. I shows the details of the hyperparameters used in our case study. Given the technical capabilities of the computer hardware employed in this work, the range of hyperparameters that could be effectively tuned was limited. The training and validation accuracy and loss over epochs are shown in Fig. 3, with an average cross-validation accuracy score of 0.803.

Parameter	Range	Selected
Epoch number	10, 20, 30	30
Batch size	32, 64	64
Activation function	relu, sigmoid, tanh	relu
Optimizer	adam, rmsprop, gd	adam
Learning rate	0.01, 0.001	0.01
Dropout ratio	0.1, 0.2, 0.3	0.1

Tab. I *Hyperparameter tuning.*

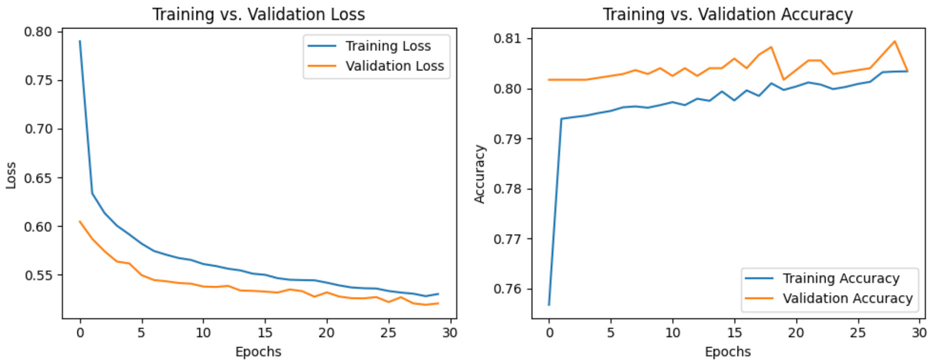

Fig. 3 *Comparison of training and validation accuracy and loss over epochs.*

Following model training, the SHAP methodology is applied for model interpretation using the DeepExplainer approach. SHAP values are computed to determine feature importance and the influence of each variable on mode choice decisions. The use of a background dataset, randomly sampled from the training data, provides a baseline for interpreting SHAP values, ensuring that feature contributions are computed in relation to representative data and, therefore, offering meaningful insights into the model's behavior. To improve computational efficiency and ensure practical feasibility on a standard laptop, the background data size was reduced to 5,000 from the original 25,972 training samples. While this reduction simplifies the analysis, the approach remains applicable to larger datasets, and results on a full sample would be expected to follow similar patterns.

Finally, SHAP plots are generated to visualize the most significant features and their contributions. These plots offer both global insights into the overall feature importance, as well as local explanations for individual predictions, giving a detailed view of the model's decision-making process. The analysis culminates in a comprehensive summary, completing the framework's interpretative process.

4.2 Data Descriptions

In this research, we used a computer-based survey, the "MaaS together" survey [19], which collected responses from 37,104 individuals across four European countries: the Czech Republic, England, Germany, and Poland. Participants were recruited through an online access panel, which ensured a diverse sample from these countries [19]. The survey collected data on the transport mode choices of the participants , and this research takes into account the following three modes: public transport, mobility as a service, and car (see Tab. II). Further details about the survey and all the variables, including latent factors, are available in the research conducted by [23].

Mode choice	Number of respondents	Percentage [%]
Public transport	29515	79.5
MaaS	2104	5.7
Car	5485	14.8

Tab. II *Mode choice distribution in the travel survey.*

As shown in Tab. II, the original dataset exhibits an imbalance, with the number of samples for each travel mode choice being significantly disproportionate (e.g., PT mode preference is 29515, whereas the Car is 5485). The imbalanced dataset can lead to biased training, as the majority class instances are used significantly more during the training process compared to the minority classes [9]. While both undersampling and oversampling can be applied to address data imbalance, these methods come with their own benefits and drawbacks. Undersampling risks losing valuable information, while oversampling may result in overfitting [9]. Given these considerations and the paper's objective, these methods are not applied in this research.

We take several explanatory variables (see Tab. III) into account in this survey, such as gender, age, and income.

Variable name	Description
Gender	Respondent's gender
Age	Respondent's age
HH_size	Number of members in the respondent's household
Income	Respondent's household income level
Residence	Size of the city where the respondent lives
Education	Respondent's education level
Profession	Respondent's occupation
Time	Time of the chosen scenario
Price	Price of the chosen scenario

Tab. III *Model input explanatory variables.*

Tab. V presents descriptive statistics for key sociodemographic variables, including household income, household size, residence, and profession. The majority of respondents report a mid-range household income (Levels 2–3), with 62% being full-time employees. Most households consist of two members (33%), while 26.3% of respondents reside in cities, and 26.6% live in large cities. These statistics provide a broad overview of the sample's sociodemographic makeup, offering valuable context for interpreting the factors that influence transportation mode choices.

Additionally, the survey includes several latent variables (see Tab. IV), as presented in [23], which are the results of a PCA run over a predefined set of questions. These variables reflect individuals' perceptions of MaaS and how it influences their decisions in general. They include the perceived usefulness of MaaS (PU) in terms of functional, emotional, social, economic, and ecological aspects, as well as the attitude toward using MaaS (ATT_fac), and other latent factors (see Tab. IV for details on the abbreviated variables).

Variable name	Description
PU_func	Functional usefulness of MaaS
PU_emot	Emotional usefulness of MaaS
PU_soc	Social usefulness of MaaS
PU_econ	Economical usefulness of MaaS
PU_eco	Ecological usefulness of MaaS
ATT_fac	Attitude toward using MaaS
FAC_share	Latent factor describing attitude towards shared economy concept
FAC_tech	Latent factor describing users technology acceptance level
FAC_env	Latent factor representing users views on environmental concerns
FAC_econ	Latent factor representing users views on economical concerns
FAC_soc	Latent factor representing social impact of surrounding opinions
FAC_safe	Latent factor representing users perception of safety importance

Tab. IV *Model input latent variables.*

4.3 Model Results and Interpretation

This section discusses the interpretability of the developed model, following the methodology outlined earlier. The outputs encompass the selected transport mode of each respondent (see Tab. II), detailing the distribution of commuters across three modes: public transport, mobility as a service (MaaS), and car. The analysis concludes with the mode choice determinants for each mode of transport.

The visualizations presented below incorporate three different interpretative tools: SHAP summary plots, beeswarm plots, and decision plots, which offer both global and local explanations. In explainable AI, local explanations use SHAP values to explain how a model makes its decisions, revealing the contributions of individual features and why the ML model arrived at its decision [11].

Variable name	Category	Percentage [%]
Household income	Level 1	6.7
	Level 2	21.9
	Level 3	40.7
	Level 4	17.8
	Level 5	10.1
	Level 6	2.8
Household size	1 person	23.7
	2 persons	33.1
	3 persons	22.9
	4 persons	15.1
	5 persons+	5.1
Residence	Town	17.3
	City	26.3
	Large city	26.6
	Megacity	19.4
	Metropolis	10.3
Profession	In training	7.3
	Full-time employee	62.5
	Part-time employee	12.8
	Senior executive	2.7
	Self-employed	6.6
	Homemaker	1.8
	Retired	3.4
	Seeking work	2.9

Tab. V *Descriptive statistics of key sociodemographic variables.*

4.3.1 Global Feature Importance

Accounting for feature interactions is a challenging task, as the individual features may have different impacts when considered in isolation compared to when they are part of a larger set of features [3]. Therefore, examining a feature's global importance is crucial for understanding its role across the entire dataset [3], which is also one of the most basic approaches to understanding a model [15], often depicted in a bar chart.

Fig. 4 presents the mean absolute SHAP values, which represent the average magnitude of each variable's impact on the model output across different transport modes. The most important factor across all types of transport seems to be how people feel about using MaaS, labeled as ATT_fac. Age also plays a significant role in shaping how people decide which mode of transport to use. In contrast, time and gender appear to be the least influential factors in determining transport mode choice, as they exhibit minimal average SHAP values across all transport modes, indicating their limited impact on the model's predictions.

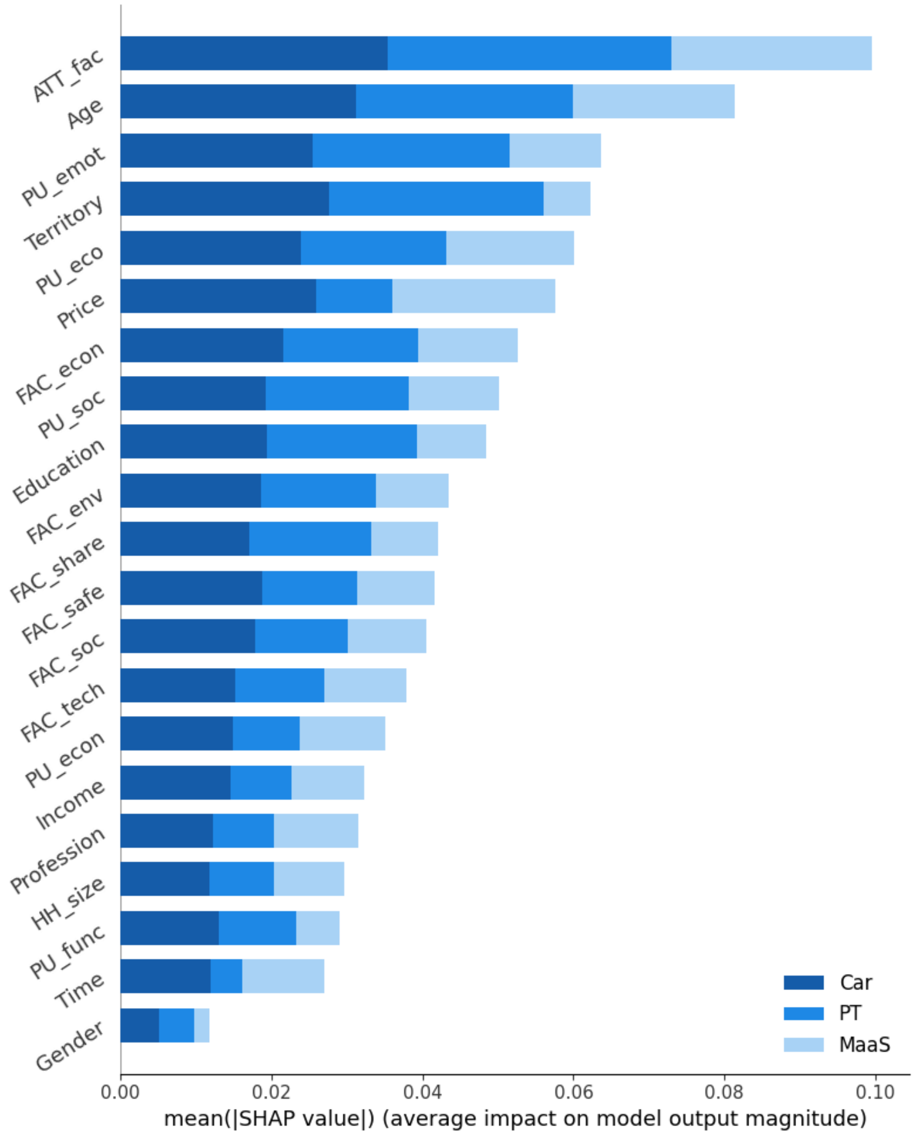

Fig. 4 *SHAP summary plot of mode choice determinants for public transport, MaaS, and car.*

4.3.2 Local Explanations Overview

Detailed summaries of the overall model and individual features can be obtained through local explanations [15]. Generating a beeswarm-style SHAP summary plot provides a comprehensive visualization that depicts the magnitude, prevalence, and direction of the effects of each feature on the final mode choice selection [15].

The top five variables influencing travel mode choices are further examined in Figs. 5 through 10. These plots provide a more nuanced understanding, showing how each feature affects the model's selection of transport modes, either positively or negatively, as depicted in the beeswarm plots.

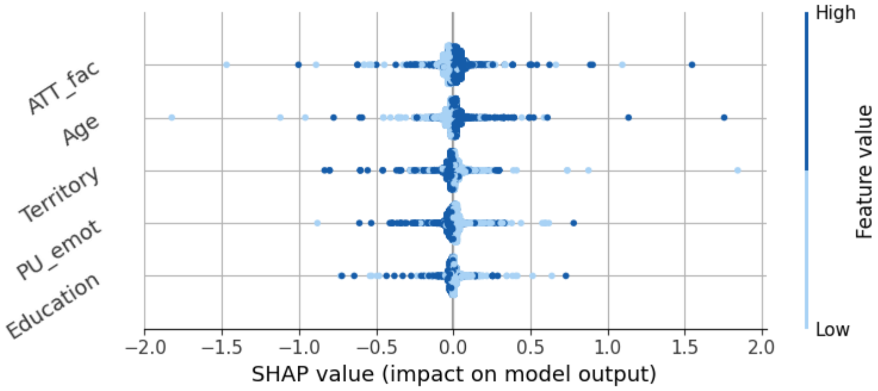

Fig. 5 *SHAP beeswarm plot of determinants for public transport mode choice.*

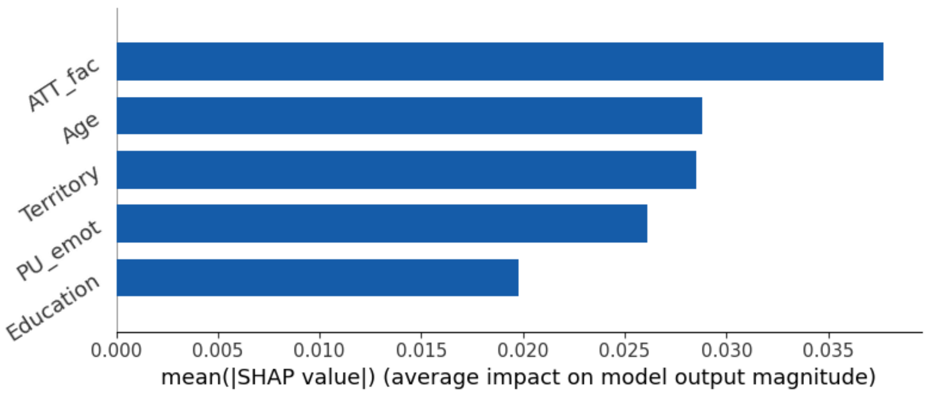

Fig. 6 *SHAP summary plot of determinants for public transport mode choice.*

In beeswarm plots, each data point represents an individual observation, with its position along the x-axis indicating the SHAP value, which reflects the extent to which a feature influences the prediction positively or negatively. Darker blue dots represent high feature values, while lighter blue dots represent low feature values, helping visualize how different feature values contribute to the model's output. For instance, in Fig. 5, high ATT_fac and age values are associated with positive SHAP values, meaning they increase the likelihood of choosing public transport. On the other hand, higher values of territory, PU_emot, and education are linked to negative SHAP values, suggesting they reduce the probability of selecting public transport.

The summary plots demonstrate that attributes such as ATT_fac and age are consistently among the most influential variables across public transport, mobility as a service, and car transport modes (see Figs. 6, 8, and 10). This consistency suggests that these factors play a pivotal role in determining transport mode choice, regardless of the specific mode. However, a more in-depth analysis of the beeswarm plots (See Figs. 5, 7, and 9) reveals notable differences in how these features influence the model's predictions, particularly in terms of the distribution and magnitude of SHAP values.

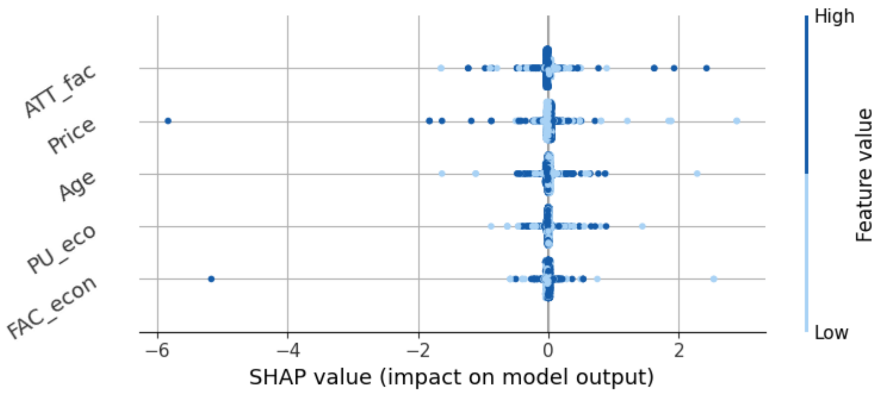

Fig. 7 *SHAP beeswarm plot of determinants for MaaS mode choice.*

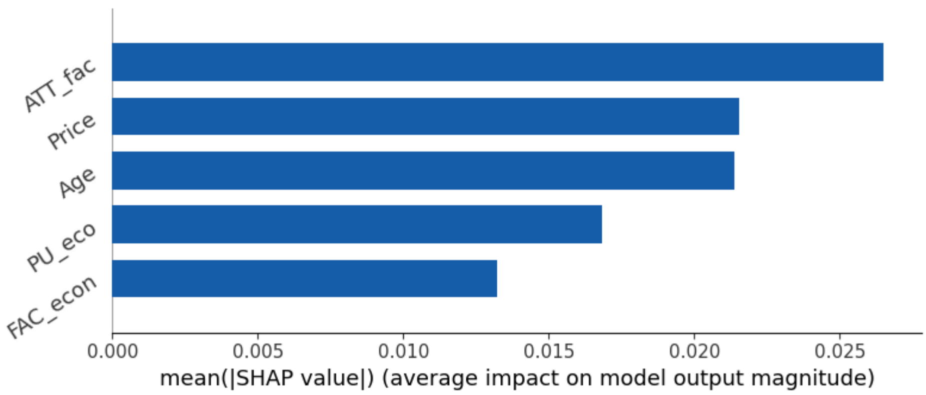

Fig. 8 *SHAP summary plot of determinants for MaaS mode choice.*

The beeswarm plot for public transport exhibits a wider range of SHAP values, with more visible data points (dots) spread across the plot. This suggests the model is capturing high variability in the feature values. This reflects diverse user behavior, where various groups of users respond differently to the same feature, resulting in a broader spectrum of predicted outcomes.

In contrast, the beeswarm plots for MaaS and Car modes show less variability. The SHAP values are more closely clustered around zero, and there are fewer extreme values represented by the darker and lighter blue dots. This suggests that while features like ATT_fac and age remain important, their influence is less pronounced in these modes. In other words, user choices in the MaaS and car modes are influenced by these factors in a more consistent way, resulting in a smaller range of variation in the model's predictions.

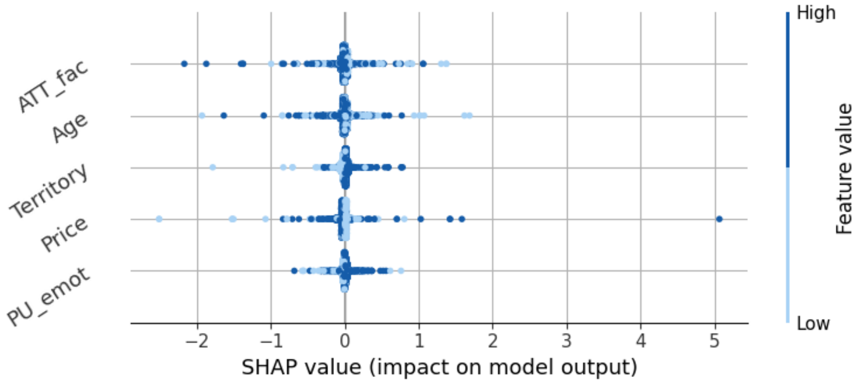

Fig. 9 *SHAP beeswarm plot of determinants for car mode choice.*

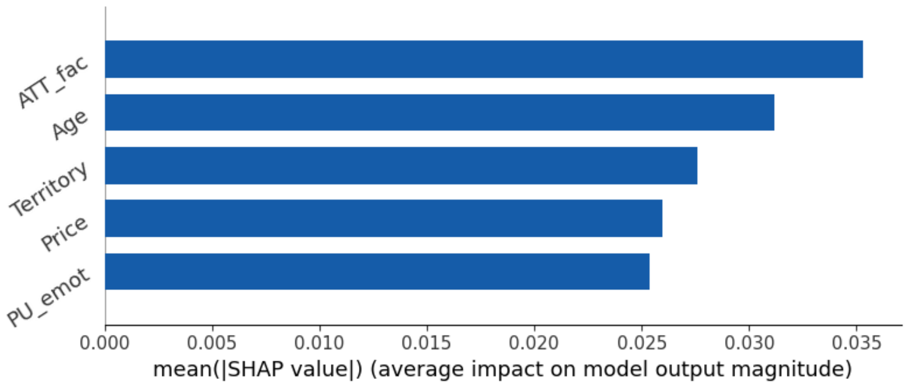

Fig. 10 *SHAP summary plot of determinants for car mode choice.*

Following the insights gained from the SHAP beeswarm plots, we now explore SHAP decision plots. Decision plots visualize how complex models make decisions by using cumulative SHAP values, with each plotted line explaining a single model prediction. Decision plots are a suitable option when there are many predictors or features in the dataset that need to be visualized [11].

The decision plots below display a random sample of 100 observations for each transport mode. This subsampling is necessary due to the large background dataset of 5,000 observations. The x-axis represents the model output value, indicating the likelihood of choosing public transport, while the y-axis lists the features in descending order of their average contribution.

The SHAP decision plot in Fig. 11 illustrates how each feature contributes to the model output for individual observations of the public transport mode. ATT_fac (attitude factor) and age stand out as the key variables, having significant positive and negative influences on the model output, depending on the specific observation. Additionally, territory and PU_emot, which represents the emotional usefulness of MaaS, play important roles and impact the usage of public transport.

The decision plot for the MaaS mode (see Fig. 12) exhibits a similar pattern, with ATT_fac once again being the most influential feature, followed by territory, age, and PU_emot. Likewise, the decision plot for the car mode (see Fig. 13) shows that ATT_fac and age continue to be the dominant feature contributions, consistent with the patterns observed for the other transport modes.

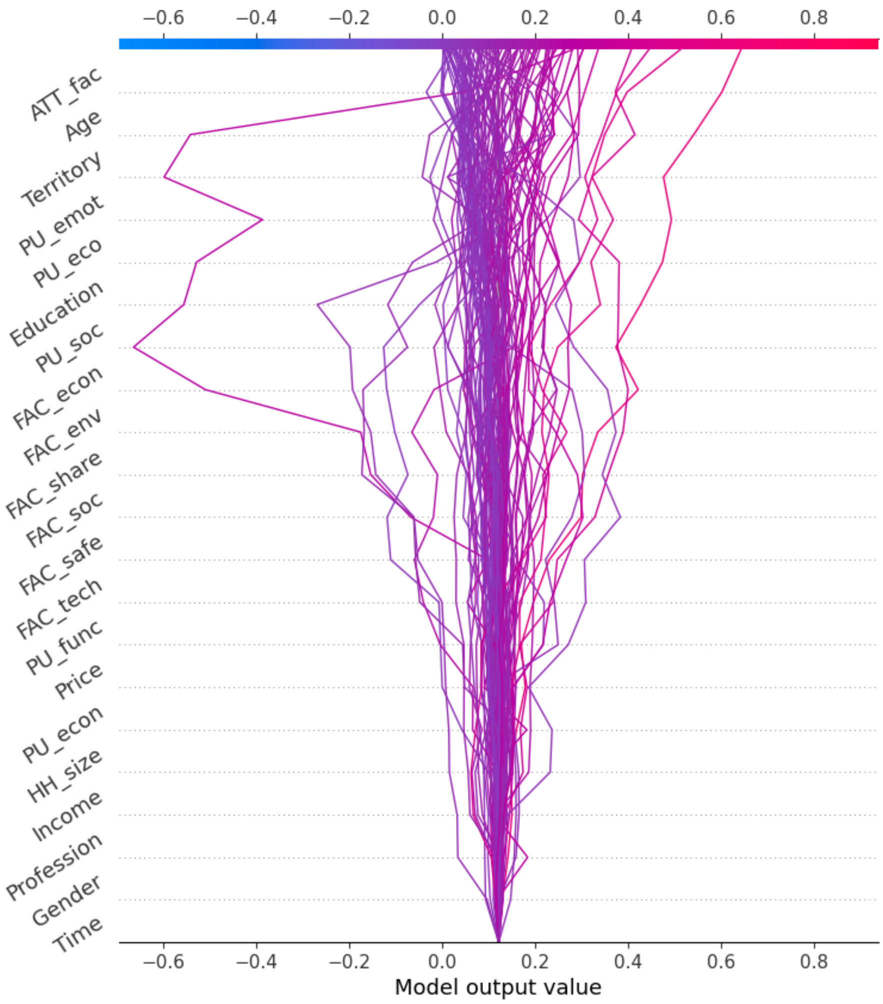

Fig. 11 *SHAP decision plot of determinants for public transport mode choice.*

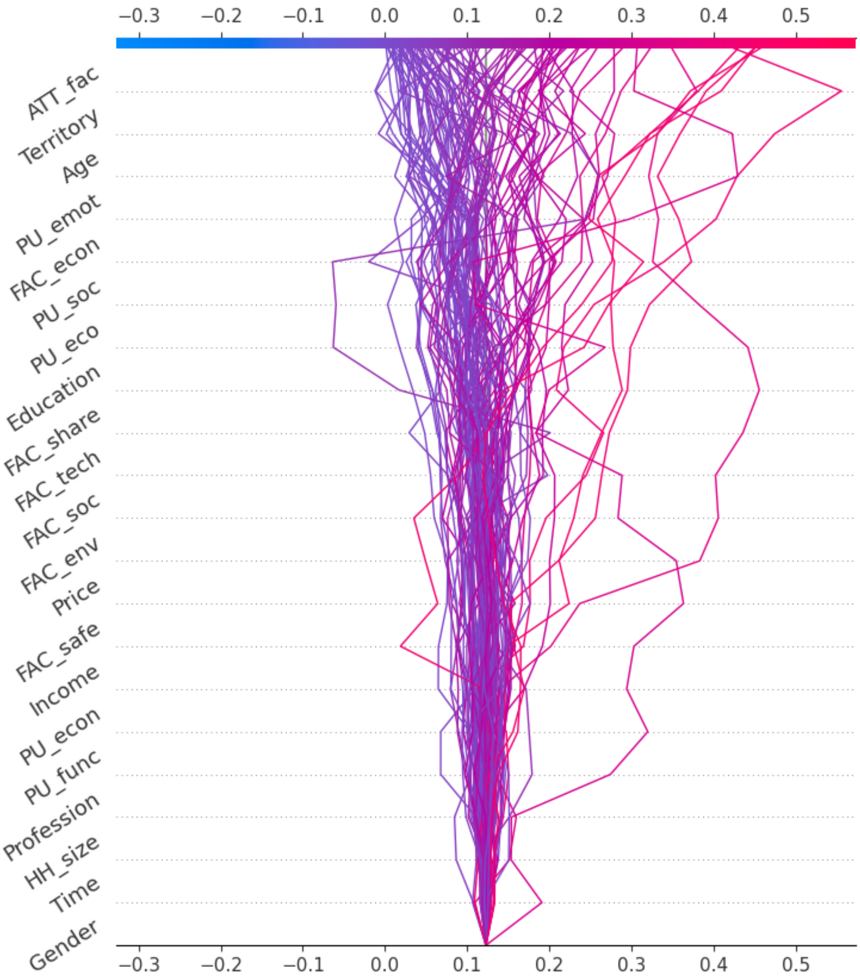

Fig. 12 *SHAP decision plot of determinants for MaaS mode choice.*

The difference in feature importance can be seen between the decision plots and the summary plots, but the reason is that these two plots serve different purposes. Summary plots show the global importance of features across the entire dataset. They aggregate the mean absolute SHAP values for each feature, providing a high-level view of which features are most influential on average for the entire population. In contrast, decision plots focus on individual predictions or selected samples. They visualize how each feature cumulatively impacts specific model outputs, which can highlight different feature contributions compared to the summary plot because the decision plot reflects local behavior for a subset of observations, not the entire dataset. This means that it focuses on the local behavior of the model, which might highlight different features than those emphasized by the global summary plot.

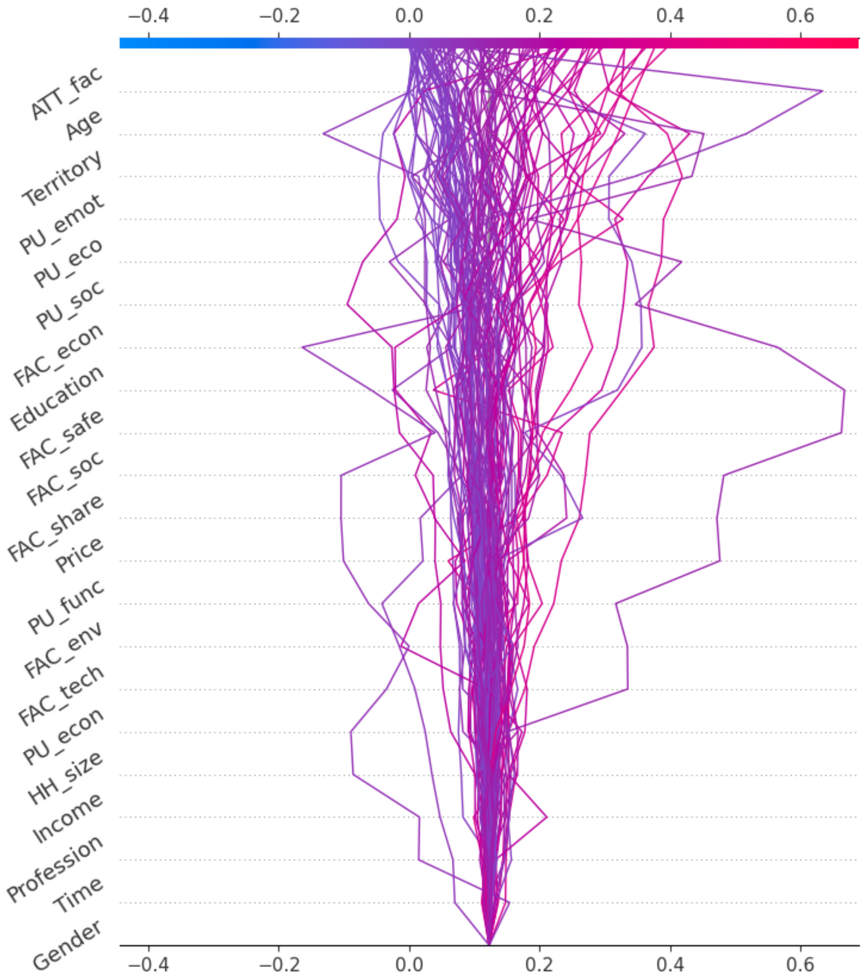

Fig. 13 *SHAP decision plot of determinants for car mode choice.*

5. Discussion

Introducing interpretable AI methods such as SHAP enhances our understanding of complex black-box models such as deep neural networks and makes these models more transparent so that we can better comprehend their decision-making processes. This approach represents a significant contribution to understanding the factors that influence mode choice decisions.

Traditional statistical models, while useful for interpreting linear relationships, often fall short when it comes to capturing the complexity and nonlinearity inherent in travel behavior. By employing a DNN and SHAP, we were able to model and understand these complex relationships, including latent interactions between socioeconomic, environmental, and behavioral factors that influence individuals' transportation choices.

Building on this methodological framework, it is crucial to investigate the specific travel mode choices available, including options such as public transport, mobility as a service (MaaS), and cars. Analyzing these travel mode choices using machine learning models can enhance our understanding and prediction of travel demands [4]. To better understand these choices, it is important to examine the factors influencing the usage of public transport, MaaS, and car modes. The existing literature has identified numerous variables that influence public transportation ridership, including factors related to urban demographics, public transport network characteristics, the availability of alternative transportation modes, and economic conditions [6].

Numerous studies examining the determinants of public transport usage have identified key factors that can be categorized into user-related aspects and system-level attributes [27]. User-related aspects include socio-demographics, such as age, income, and education, which significantly affect individuals' likelihood of using transit, along with accessibility to transit stops and the characteristics of the built environment [27]. The socio-demographic factors identified in our study, particularly age and education, are also significant, with age being one of the most influential factors across all transport modes. This finding is consistent with the conclusions of other research in this field.

The literature review further revealed that key determinants of individual travel mode choice include the prices of public transport and private car transport [6]. Prior studies have also shown that system-level attributes, such as level of service, pricing structures, and external influences like economic conditions, car ownership rates, and parking policies, play crucial roles in shaping public transportation ridership [27]. Additionally, consistent with these findings, our results indicate that price is also a significant feature, particularly for the MaaS and car modes.

6. Conclusions

This study presents a novel approach to analyzing individual travel behavior through the integration of Deep Neural Networks (DNN) and SHAP, an interpretable AI technique, to uncover the factors driving mode choice decisions. While DNNs are powerful in terms of predictive performance and usually outperform traditional statistical approaches, their "black-box" nature limits their usability for policy-making and trust-building. SHAP addresses this challenge by providing clear, interpretable insights into the contribution of each variable, offering a transparent way to understand how specific features—such as income, gender, distance to destination, and environmental conditions—influence travel decisions. This interpretability is crucial not only for researchers but also for transportation planners and policymakers who need actionable insights to design targeted interventions.

The use of DNN and SHAP offers several advantages. First, the ability of DNNs to model nonlinear and complex relationships makes them highly suitable for transportation data, where individual decisions are influenced by a myriad of interrelated factors. Second, SHAP enables us to break down the model's decision-making process, ensuring transparency and interpretability, which is critical for stakeholders who need to trust and act on the results. Third, the findings provide actionable insights that can help in crafting transportation policies that promote sustainable mode choices, optimize public transport, and influence behavior towards more eco-friendly alternatives. Lastly, this approach can be scaled to larger datasets and adapted to other regions or contexts, making it versatile for future research or real-world applications.

In conclusion, our study highlights the significant potential of combining deep neural network models with interpretable AI methods to not only improve the accuracy of travel behavior models but also make them accessible and practical for real-world application. By understanding the factors that shape mode choice in a transparent way, transportation planners can develop more informed policies that are aligned with the goals of sustainability, such as promoting environmentally friendly transportation options as well as efficiency. Future research should aim to expand on this work by refining these methods to further enhance their applicability and comparing the performance and interpretability of this approach with other mode choice modeling techniques. This comparative analysis could provide valuable insights into the strengths and limitations of different modeling approaches.

Credit Author Statement

Halil Çevik: Conceptualization, methodology, software, formal analysis, writing – original draft, writing – review & editing, visualization.

Ondřej Přibyl: Writing – original draft, writing – review & editing, supervision.

Shoaib Samandar: Writing – review & editing.

Acknowledgement

This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS23/129/OHK2/2T/16.

References

[1] CHEN L., TSAO Y., SHEU J.-T. Using deep learning and explainable artificial intelligence in patients' choices of hospital levels. *arXiv preprint arXiv:2006.13427*. 2020, doi: 10.48550/arXiv.2006.13427.

[2] CHENG L., CHEN X., DE VOS J., LAI X., WITLOX F. Applying a random forest method approach to model travel mode choice behavior. *Travel behaviour and society*. 2019, 14, pp. 1–10, doi: 10.1016/j.tbs.2018.09.002.

[3] COVERT I., LUNDBERG S.M., LEE S.-I. Understanding global feature contributions with additive importance measures. *Advances in Neural Information Processing Systems*. 2020, 33, pp. 17212–17223, doi: 10.48550/arXiv.2004.00668.

[4] HAGENAUER J., HELBICH M. A comparative study of machine learning classifiers for modeling travel mode choice. *Expert Systems with Applications*. 2017, 78, pp. 273–282, doi: 10.1016/j.eswa.2017.01.057.

[5] HATAMI F., RAHMAN M.M., NIKPARVAR B., THILL J.-C. Non-linear associations between the urban built environment and commuting modal split: A random forest approach and SHAP evaluation. *IEEE Access*. 2023, 11, pp. 12649–12662, doi: 10.1109/access.2023.3241627.

[6] INGVARDSON J.B., NIELSEN O.A. How urban density, network topology and socio-economy influence public transport ridership: Empirical evidence from 48 European metropolitan areas. *Journal of Transport Geography*. 2018, 72, pp. 50–63, doi: 10.1016/j.jtrangeo.2018.07.002.

[7] IWASAKI Y., SAWADA R., STANEV V., ISHIDA M., KIRIHARA A., OMORI Y., SOMEYA H., TAKEUCHI I., SAITOH E., YOROZU S. Identification of advanced spin-driven thermoelectric materials via interpretable machine learning. *npj Computational Materials*. 2019, 5(1), pp. 103, doi: 10.1038/s41524-019-0241-9.

[8] JAHANGIRI A., RAKHA H.A. Applying machine learning techniques to transportation mode recognition using mobile phone sensor data. *IEEE transactions on intelligent transportation systems*. 2015, 16(5), pp. 2406–2417, doi: 10.1109/tits.2015.2405759.

[9] KASHIFI M.T., JAMAL A., KASHEFI M.S., ALMOSHAOGEH M., RAHMAN S.M. Predicting the travel mode choice with interpretable machine learning techniques: A comparative study. *Travel Behaviour and Society*. 2022, 29, pp. 279–296, doi: 10.1016/j.tbs.2022.07.003.

[10] KIM E.-J. Analysis of travel mode choice in Seoul using an interpretable machine learning approach. *Journal of Advanced Transportation*. 2021, 2021(1), pp. 6685004, doi: 10.1155/2021/6685004.

[11] LE T.-T.-H., KIM H., KANG H., KIM H. Classification and explanation for intrusion detection system based on ensemble trees and SHAP method. *Sensors*. 2022, 22(3), pp. 1154, doi: 10.3390/s22031154.

[12] LI X., XIONG H., LI X., WU X., ZHANG X., LIU J., BIAN J., DOU D. Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond. *Knowledge and Information Systems*. 2022, 64(12), pp. 3197–3234, doi: 10.1007/s10115-022-01756-8.

[13] LI Z. Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost. *Computers, Environment and Urban Systems*. 2022, 96, pp. 101845, doi: 10.1016/j.compenvurbsys.2022.101845.

[14] LIU M., NING Y., YUAN H., ONG M.E.H., LIU N. Balanced background and explanation data are needed in explaining deep learning models with SHAP: An empirical study on clinical decision making. *arXiv preprint arXiv:2206.04050*. 2022, doi: 10.48550/arXiv.2206.04050.

[15] LUNDBERG S.M., ERION G., CHEN H., DEGRAVE A., PRUTKIN J.M., NAIR B., KATZ R., HIMMELFARB J., BANSAL N., LEE S.-I. Explainable AI for trees: From local explanations to global understanding. *arXiv preprint arXiv:1905.04610*. 2019, doi: 10.48550/arXiv.1905.04610.

[16] LUNDBERG S.M., ERION G.G., LEE S.-I. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*. 2018, doi: 10.48550/arXiv.1802.03888.

[17] LUNDBERG S.M., LEE S.-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems*. 2017, 30, doi: 10.48550/arXiv.1705.07874.

[18] MARTÍN-BAOS J.Á., LÓPEZ-GÓMEZ J.A., RODRIGUEZ-BENITEZ L., HILLEL T., GARCÍA-RÓDENAS R. A prediction and behavioural analysis of machine learning methods for modelling travel mode choice. *Transportation research part C: emerging technologies*. 2023, 156, pp. 104318, doi: 10.1016/j.trc.2023.104318.

[19] MATOWICKI M., AMORIM M., KERN M., PECHERKOVA P., MOTZER N., PRIBYL O. Understanding the potential of MaaS–An European survey on attitudes. *Travel Behaviour and Society*. 2022, 27, pp. 204–215, doi: 10.1016/j.tbs.2022.01.009.

[20] MENG M., MEMON A.A., WONG Y.D., LAM S.-H. Impact of traveller information on mode choice behaviour. In: *Proceedings of the Institution of Civil Engineers-Transport*, 1, 2018, pp. 11–19. doi: 10.1680/jtran.16.00058.

[21] MONTAVON G., SAMEK W., MÜLLER K.-R. Methods for interpreting and understanding deep neural networks. *Digital signal processing*. 2018, 73, pp. 1–15, doi: 10.1016/j.dsp.2017.10.011.

[22] MOONS E., WETS G., AERTS M. Nonlinear Models for Determining Mode Choice: (Accuracy Is Not Always the Optimal Goal). In: *Progress in Artificial Intelligence: 13th Portuguese Conference on Aritficial Intelligence, EPIA 2007, Workshops: GAIW, AIASTS, ALEA, AMITA, BAOSW, BI, CMBSB, IROBOT, MASTA, STCS, and TEMA, Guimarães, Portugal, December 3-7, 2007. Proceedings 13*, 2007, pp. 183–194. doi: 10.1007/978-3-540-77002-2_16.

[23] MOTZER N., SOARES AMORIM M.R., MATOWICKI M., KERN M., PECHERKOVÁ P. Exploring various dimensions of perceived usefulness on the intention to use Mobility-as-a-Service. *Human Factors in Transportation*. 2023, 95(95).

[24] NAM D., CHO J. Deep neural network design for modeling individual-level travel mode choice behavior. *Sustainability*. 2020, 12, pp. 7481, doi: 10.3390/su12187481.

[25] PRENDIN F., PAVAN J., CAPPON G., DEL FAVERO S., SPARACINO G., FACCHINETTI A. The importance of interpreting machine learning models for blood glucose prediction in diabetes: an analysis using SHAP. *Scientific reports*. 2023, 13(1), pp. 16865, doi: 10.1038/s41598-023-44155-x.

[26] QIN H., GAO J., WU Y.-J., YAN H. Analysis on context change and repetitive travel mode choices based on a dynamic, computational model. *Transport Policy*. 2019, 79, pp. 155–164, doi: 10.1016/j.tranpol.2019.04.003.

[27] SHAABAN K., SIAM A. Review of factors affecting public transportation ridership. In: *2022 Intermountain Engineering, Technology and Computing (IETC)*, 2022, pp. 1–5. doi: 10.1109/ietc54973.2022.9796772.

[28] SHRIKUMAR A., GREENSIDE P., KUNDAJE A. Learning important features through propagating activation differences. In: *International conference on machine learning*, 2017, pp. 3145–3153. doi: 10.48550/arXiv.1704.02685.

[29] ULLAH I., LIU K., YAMAMOTO T., ZAHID M., JAMAL A. Modeling of machine learning with SHAP approach for electric vehicle charging station choice behavior prediction. *Travel Behaviour and Society*. 2023, 31, pp. 78–92, doi: 10.1016/j.tbs.2022.11.006.

[30] WANG D., THUNÉLL S., LINDBERG U., JIANG L., TRYGG J., TYSKLIND M. Towards better process management in wastewater treatment plants: Process analytics based on SHAP values for tree-based machine learning methods. *Journal of Environmental Management*. 2022, 301, pp. 113941, doi: 10.1016/j.jenvman.2021.113941.

[31] WANG S., WANG Q., BAILEY N., ZHAO J. Deep neural networks for choice analysis: A statistical learning theory perspective. *Transportation Research Part B: Methodological*. 2021, 148, pp. 60–81, doi: 10.1016/j.trb.2021.03.011.

[32] YUAN H., LIU M., KANG L., MIAO C., WU Y. An empirical study of the effect of background data size on the stability of SHapley Additive exPlanations (SHAP) for deep learning models. *arXiv preprint arXiv:2204.11351*. 2022, doi: 10.48550/arXiv.2204.11351.

[33] ZHANG Z., JI C., WANG Y., YANG Y. A customized deep neural network approach to investigate travel mode choice with interpretable utility information. *Journal of Advanced Transportation*. 2020, 2020(1), pp. 5364252, doi: 10.1155/2020/5364252.