Transportation Research Part A 173 (2023) 103727

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

Comparing and contrasting choice model and machine learning techniques in the context of vehicle ownership decisions

Azam Ali, Arash Kalatian, Charisma F. Choudhury *

Institute for Transport Studies, University of Leeds, UK

A R T I C L E I N F O

Keywords:
Choice modelling
Machine learning
Explainable machine learning
Vehicle ownership model
Developing countries
Car ownership

A B S T R A C T

In recent years, planners have started considering Machine Learning (ML) techniques as an alternative to discrete choice models (CM). ML techniques are primarily data-driven and typically achieve better prediction accuracy compared to CM. However, it is hypothesized that since the ML techniques do not have the strong grounding to economic theory as the CMs, they may not perform well in contexts that are radically different from the 'training' scenario. It is also hypothesized that the relative prediction performance may be affected by the metrics used for comparing the models.

This research aims to test these two hypotheses empirically by modelling vehicle ownership choices using household survey data from Dhaka, Bangladesh collected in 2004, 2010 and 2019. The performances of CM (multinomial logit) and ML techniques (neural networks and gradient boosting trees) have been compared using log-likelihood and mean absolute percentage error of market shares.

The results indicate that the multinomial logit model (MNL) with a piecewise linear transformation of the household income, has the best performance in terms of log-likelihood and mean absolute percentage error of market shares. This is followed by Neural Networks (NN) and Gradient Boosting Trees (GBT). The results thus provide empirical evidence that the ML techniques do not consistently outperform CM. Moreover, the difference in the performance of the models further increases if the prediction scenario is substantially different. This reinforces the hypothesis that CMs, with their behavioural underpinning, are better suited for long-term forecasting than data-driven ML approaches, especially if the population and network attributes are expected to change substantially. These findings will be useful for planners and policy makers in the selection of the appropriate tool for forecasting travel demand.

1. Introduction

Travel behaviour models have historically relied on Choice Models (CM) based on theories of economics and psychology. However, the availability of large datasets on human mobility in recent years has led to an increased interest in deploying Machine Learning (ML) techniques to predict travel behaviour. These ML techniques use parametric and nonparametric algorithms to 'learn' directly from the data (Van Cranenburgh et al., 2022; Walker et al., 2019). Such data-driven learning enables researchers to model large and complex datasets without making explicit assumptions about the behavioural motivations or the relationship between the dependent and the

* Corresponding author.
E-mail addresses: ts19aa@leeds.ac.uk (A. Ali), arash.kalatian@gmail.com (A. Kalatian), C.F.Choudhury@leeds.ac.uk (C.F. Choudhury).

https://doi.org/10.1016/j.tra.2023.103727

Available online 6 June 2023

independent variables. However, the lack of behavioural underpinning also poses the risk of overfitting the data, which can lead to unsatisfactory performance in predicting behaviour in scenarios that are radically different from the current situation. This is particularly an issue of concern in terms of the spatial and temporal transferability of the ML models. Further, outputs of the ML models are inherently uninterpretable and typically do not allow modellers to reliably extract information such as the value of time and willingness-to-pay for cost-benefit analyses (Aboutaleb et al., 2021).

In recent years, there have been a growing number of studies where researchers have been interested in finding if ML techniques can indeed outperform CM in terms of achieving higher prediction accuracy. The majority of these studies are concerned with mode choices (e.g. Hagenauer and Helbich, 2017; Hillel et al., 2020; Wang et al., 2020a, 2020b; Zhao et al., 2020 etc.), while a few studies have focused on route choice modelling (Yao and Bekhor, 2020) and vehicle ownership decisions (Basu, 2019; Paredes et al., 2017). Several of these studies (Hagenauer and Helbich, 2017; Lee et al., 2018; Lu et al., 2021; Paredes et al., 2017; Salas et al., 2022) have found that ML models outperform CM in prediction accuracy. However, a recent review of nearly seventy research studies by Hillel et al. (2020) reports that there are serious technical issues with the majority of these studies. The majority of studies examined in the review by Hillel et al. (2020) compare CM and ML techniques using prediction metrics (such as overall accuracy and/or choice specific accuracy), thereby treating these models as deterministic and ignoring the probabilistic nature of the analytical models. Minimal focus has been given to evaluation metrics based on probabilities, especially log-likelihood, arguably the most relevant metric for evaluating probabilistic discrete choices. Although prediction accuracy in disaggregate level is of core interest in some marketing and consumer choice applications (Basu, 2019); in the context of travel behaviour, the aggregate demands (deduced from disaggregate choices) are generally of more interest (Ben-Akiva and Lerman, 1985; Train, 2009). Using predictions rather than probabilities to calculate market shares (as typically done in case of ML), can lead to issues like aggregation bias and/or class imbalance (Hillel et al., 2018, 2020; Walker et al., 2019). This raises questions about the role of the performance metrics used on the outcome of the comparison between ML and CM.

Another issue of practical interest is the comparison of the planning and policy insights gained from the ML and CM. To explain the choices predicted by ML models, visualization techniques such as feature importance, partial dependence plots and individual conditional expectation plots, and computation of elasticities and prototypical examples have been used in the literature (Alwosheel et al., 2019; Wang et al., 2020a, 2020b; Zhao et al., 2019, 2020). Although there has been limited success due to ML techniques' lack of behavioural underpinning, these studies recommend researchers to use new techniques to explain ML models and search for better regularization approaches. Furthermore, other recent ML explainability techniques, e.g. Local Interpretable Model-agnostic Explanations (LIME), (Ribeiro et al., 2016) and Shapley Additive Explanations (SHAP), (Lundberg and Lee, 2017), have not been adequately explored in the context of interpreting travel behaviour. For instance, the application of SHAP in the context of transport research primarily includes analyzing traffic accidents (Parsa et al., 2020; Wen et al., 2021; Yang et al., 2021), analyzing travel patterns of people with reduced mobility (Pineda-Jaramillo, 2021) and explaining the pedestrian wait time before crossing (Kalatian and Farooq, 2021). These emerging methods of 'explainable machine learning' provides the opportunity to compare the planning and policy implications of the outputs derived from ML and CM.

Motivated by the research gaps, the study compares ML and CM in the context of vehicle ownership decisions in a developing country using household survey data from three different years. The specific questions the research aims to answer are:

(1) How do the performances of CM and ML vary in terms of measures of performance such as log-likelihood?
(2) What is the relative performance of ML techniques and CMs when there are substantial differences in the application contexts?
(3) How similar or dissimilar are the outputs and the insights gained for planning applications and policy formulation?

Household survey data collected in Dhaka, Bangladesh in 2005, 2010, and 2019 have been used in this regard. Household survey data collected in 2010 have been used to estimate a vehicle ownership model[1]. The estimated model is then used to backcast and forecast vehicle ownership in 2005 and 2019, respectively. Probabilistic metrics are used to compare the performance of the models. Partial Dependence Plots and SHAP values are used to explain the ML models, and the outputs are then compared to the relative importance of attributes derived from the CM.

The rest of the paper is organized as follows: the background theory is presented first, followed by descriptions of the datasets and models. The results are presented next, along with the comparison of CM and ML Models. The findings are summarized in the concluding section.

2. Background theory

The background theories of the CM and ML techniques are summarised briefly in the following sections.

2.1. Choice models

Disaggregate vehicle ownership models are typically modelled using unordered or ordered choice models (Anowar et al., 2014; de Jong et al., 2004). Ordered choice models, like ordered logit (OL) and ordered probit (OP), are based on the hypothesis that vehicle

[1] The 2010 data was chosen as the estimation and training dataset as it is much larger than the 2005 dataset (see Section 4 for details).

ownership decisions are ordered decisions, i.e. households will progress from owning no vehicles to one vehicle, then two, and so on. Similarly, the decision to own a vehicle is based on a single continuous latent variable which is divided into different thresholds or partitions that demarcate the different discrete outcomes (Bhat and Pulugurta, 1998). However, ordered models are quite a simplification for vehicle decisions in case of multiple types of vehicle ownership decisions, as a decision-maker may attribute different utilities to different vehicle ownership choices (rather than a continuous one for all the choices). Further, it is problematic to arrange the choices in an order when there are different types of vehicles, e.g. motorcycles and cars (Bhat and Pulugurta, 1998). Hence, unordered choice models like multinomial logit (MNL) and nested logit (NL) models are typically preferred over ordered choice models in the context of multiple types of vehicle mixes.

2.1.1. Multinomial logit models (MNL)

The basis of unordered choice models is the utility maximization theory, where each decision maker assigns a utility to each alternative in the choice set (J) and chooses the alternative with the highest utility. However, as utility is not fully observable to the modeller, a probabilistic approach is considered, which leads to the formation of random utility theory (see Train, 2009 for details). A decision-maker, n associates a utility $U_{n,i}$ to the alternative i which equals to the sum of a deterministic or observable utility $V_{n,i}$ and a random error term $\varepsilon_{n,i}$. Assuming the error terms are distributed independently and identically, using a type 1 extreme value distribution, the probability P of decision-maker n choosing alternative $i$can be expressed as follows:

$$P_{n,i} = \frac{e^{V_{n,i}}}{\sum_{j=1}^{J} e^{V_{n,j}}} \text{where} V_{n,i} = f\left(\beta, x_{n,i}\right) \tag{1}$$

Where $x_{n,i}$ denotes the variables influencing the utility of individual n for alternative i and β denotes the model parameters. The β parameters are determined by maximizing the following log-likelihood (LL) function where y is 1 if the decision-maker n chooses the alternative i:

$$LL(\beta) = \sum_{n=1}^{N} \sum_{i} y_{n,i} \ln\left(P_{n,i}\right) \tag{2}$$

Nested logit (NL) models relax the requirement of the error terms to be independently and identically distributed and allow for correlation among the error terms of similar alternatives. As a previous study on vehicle ownership models using the 2005 and 2010 datasets show that the data does not support NL models (see Flavia and Choudhury, 2019 for details), they are not elaborated in this section.

2.2. Machine learning algorithms

For this study, neural networks and ensembles of decision trees, two of the most widely used and the best performing ML algorithms in the context of travel behaviour modelling (Hillel et al., 2020), are selected to be evaluated in this study. A detailed review of the studies comparing CM and ML techniques is included in Appendix I.

2.2.1. Neural network

Neural networks can be described as a set of linear equations with weights and biases connecting at 'neurons'. Each neuron is transformed by a nonlinear activation function as mentioned in equation (3) (Bishop, 2006). The weight, W and biases b of the neural network are adjusted using maximum likelihood principle which minimizes the cross entropy or maximizes the average log-likelihood of the function (same as equation (2)) (Goodfellow et al., 2017). The activation function f. used in the last layer of a neural network trained for a multi-class classification problem is a soft-max function (equation (4)), which normalizes and calibrates the output of the hidden layers between 0 and 1 and ensures that the sum of the choice probabilities is 1 (Bishop, 2006; Goodfellow et al., 2017).

$$Z_i = f.(W^{*}x + b) \tag{3}$$

$$P(Class\ i \mid x) = \frac{e^{Z_i}}{\sum_i e^{Z_i}} \tag{4}$$

2.2.2. Decision tree and ensemble learning

The decision tree is a popular and easy to follow supervised Machine Learning algorithm which splits the dataset into different classes based on simple 'if-else' statements (Hastie et al., 2009). However, there is a high variance in the building of decision trees. Similarly, decision trees only allow discrete outcomes, i.e. 1 or 0 for a binary outcome, rather than choice probabilities (Hastie et al., 2009; Hillel et al., 2018). Ensemble Learning helps to mitigate the stability issues in decision trees and allows the calculation of choice probabilities (Hillel et al., 2018). The most widely used and robust ensemble learning boosting algorithm is Gradient Boosting Tree, where a gradient descent method is used (see Hastie et al. (2009) for details).

2.3. Challenges of training ML models

Even though there is no need to specify the utility function in ML techniques, as in the case of choice models, a significant amount of effort needs to be deployed into training the ML techniques.

All ML models suffer from bias-variance trade-offs or the balance between under and overfitting (Goodfellow et al., 2017). If the ML model has low complexity, the model would not fully capture the data generating process. On the other hand, in high levels of complexity, the model may be trained on noise or errors, leading to low performance on unseen datasets and resulting overfitting (Goodfellow et al., 2017). The performance of all ML techniques, especially NN, suffers from three main challenges, i.e. "high sensitivity to hyperparameters, model non-identification, and local irregularity" (Wang et al., 2020a).

Hyperparameters are different tools that control the learning behaviour of ML techniques (Goodfellow et al., 2017). For the selection of the hyperparameters, *k* fold cross validation is recommended to reduce the risks of overfitting. It is recommended to carry out numerous trainings, for instance Wang et al. (2020a) and Wang and Ross (2018) conducted 100 trainings under different initialization settings to reduce the risk of model non-identification and achieve more reliable results. Local irregularity can further be mitigated by using economic information aggregated over a population rather than using individual information and by having a large sample size (approximately 10^4) (Wang et al., 2020a).

2.4. Explainable ML techniques

In recent years, there has been a rise of explainable[2] ML techniques among the ML community to build the trust of decision-makers and modellers in using ML models for planning applications (Doshi-Velez and Kim, 2017; Molnar, 2019). These motivations have led to the formulation of techniques such as partial dependence plots or SHAP values. These techniques are model agnostic, i.e. can be used for any ML technique and are applied after the training of the ML model.

2.4.1. Partial dependence plots

First introduced by Friedman (2001), Partial dependence plots (PDP) provide a visual method to relate the dependence of an explanatory variable to the model's target, i.e. choice probability in a supervised classification problem. Partial dependence plots can also be visualized for two variables, making it possible to capture the interaction of two different variables on the choice probabilities. However, it is not possible to visualize the dependence for more than two variables. Another drawback of the partial dependence plots is that they cannot capture the presence of heterogeneity in the choices.

To calculate the partial dependence function $(\widehat{f}_s)$ of variable *s*, we first construct datasets of varying values of x_s ,while keeping the other variables constant. The trained ML model, $\widehat{f}$ is used to predict the response, i.e. choice probabilities, on the varying datasets $\widehat{f}_s$. Finally, average predicted response, is plotted against the varying values of x_s Goldstein et al., (2015). Mathematically, it is expressed in equation (5) where *N*is the total number of observations.

$$\widehat{f}_s = \frac{1}{N}\sum_{n=1}^{N}\widehat{f}\left(x_s, x_c^n\right) \tag{5}$$

2.4.2. SHAP values

SHAP is a recently popular explainable ML technique that is based on Shapley values, a game theory based approach. This technique unifies different types of explainable ML techniques to give an explanation that is locally accurate and consistent (Lundberg and Lee, 2017). The strength of SHAP is that it represents Shapley values as a linear additive model. Readers interested in completing the background theory and computation of SHAP values are referred to in the study by Lundberg and Lee (2017).

The basic concept of SHAP values is to split ML model's predicted individual choice probability into a linear additive model as mentioned in equation (6) where *M* is the number of total explanatory variables. The first step in the computation of the SHAP value is to calculate the expected or base value of the trained ML model, $\varnothing_0$ which is the choice probability when all explanatory variables are missing. Then, the effect of each variable on individual choice probability is observed by running various permutation and coalitions. For e.g., first the variable x_1 is "turned on" or activated and the corresponding change in probability is noted, then only variable x_2 is turned on and its effect is noted, then the effect of a coalition of variables x_1 and x_2 is noted. After noting the effect of different permutations and coalitions, Shapley value, which is normally used in coalitional game theory, is calculated to give a fair and unique solution to the contribution of each explanatory variable in the choice probability. For estimation of these SHAP values some approximations are used such as the independence between the explanatory variables (Lundberg and Lee, 2017).

$$f(x) = \varnothing_0 + \sum_{i=1}^{M}\varnothing_i x_i \tag{6}$$

Since SHAP values are calculated independently for each observation, this technique can uncover the heterogeneity in the choices modelled. Individual SHAP values for each attribute can further be averaged to make a feature importance plot that shows the relative importance of each explanatory variable on the choice probabilities. However, the calculation of SHAP values require high computation time and may not always be intuitive (Molnar, 2019).

[2] The terms Interpretable and Explainable ML techniques are sometimes used interchangeably in the literature. However, for conformity the term explainable ML techniques has been used in this paper. It may be noted that the explainable ML do not necessarily describe any cause-effect relationships between the variables based on an economic theory.

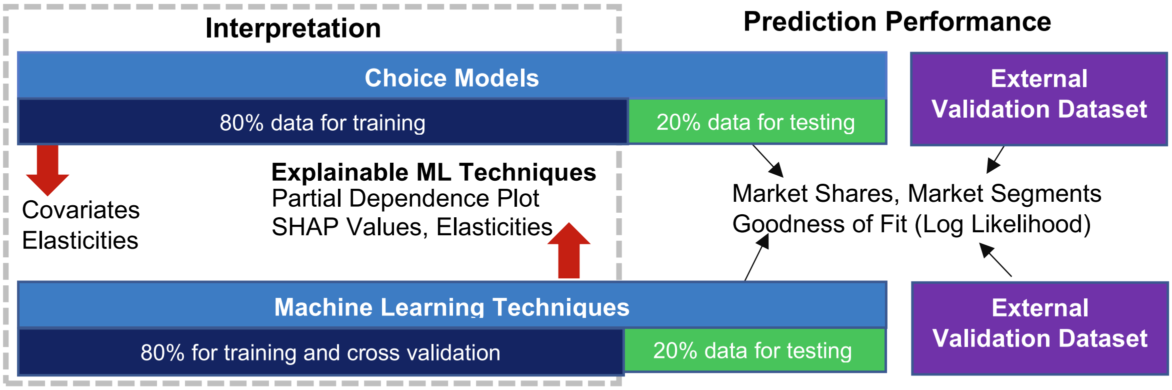

Fig. 1. Methodology to compare choice models and ML techniques.

2.4.3. Elasticities

As most ML algorithms can calculate probabilities, it is possible to calculate elasticities by inducing changes in variables similar to calculating elasticities in choice models. Aggregate elasticities for continuous variables can be approximated by using equation (7) for a 10% change in variable *i*. Individual arc elasticities can be computed using equation (8) and converted to aggregate elasticity by the probability-weighted sample enumeration method (see Hensher et al., 2005 for details).

$$E_i \approx \frac{\log\left(\frac{futuredemand_i}{basedemand}\right)}{\log(1.1)} \tag{7}$$

$$E_i = \frac{P(Yx_i = j) - P(Y|x_i = j+1)}{0.5^*[P(Yx_i = j) + P(Yx_i = j+1)]} \tag{8}$$

2.5. Comparison of CM and ML techniques

As shown in Fig. 1**,** the comparison of CM and ML techniques for this study is divided into two parts: prediction performance and interpretation**.**

For evaluating the prediction performance, CM models typically use log-likelihood (LL) and other metrics that are functions of the LL and the number of parameters (such as adjusted rho square, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)). Since the AIC, BIC, and adjusted rho square penalize models with higher number of parameters (as is typically the case of ML), the log-likelihood has been chosen as the primary criterion to compare the CM and ML models in this research. This ensures the ML models are not additionally penalized due to their dependency on large number of parameters.

While the scope of the previous studies in this area has been limited to comparing the performance of the models on the training and testing using 'hold-out' samples, the developed models can also be compared with an external dataset. The purpose of the external validation dataset is to check if the model can accurately forecast choices using data collected by a different methodology or in a different time or location (Hillel et al., 2020). One way to check and improve model specification in choice modelling is to observe if the model can retain market shares (calculated using probabilities) in different market segments based on different socio-economic features (Hess and Palma, 2019). Similar tests can also be carried out in ML models.

Choice models are simple to analyze and interpret. The β coefficients and the corresponding standard errors estimate the relative importance of attributes and allow for the calculation of economic information such as elasticity, willingness to pay, value of time. Most ML models, on the other hand, are considered as black boxes, meaning that it is much more difficult to interpret or explain them as they have high levels of nonlinearities in their structures. Visual explanation techniques such as partial dependence plots can be compared to the substitution pattern of alternatives observed in CMs. Similarly, elasticities and marginal effects can be used to quantify the effect of explanatory variables, which can be used for planning and policy formulation.

3. Data description

The datasets used in this study have been collected from three household surveys carried out in Dhaka, Bangladesh. The first household survey was carried out in 2004 (published in 2005) for the development of the Strategic Transport Plan (STP) for Dhaka (World Bank, 2005). The second survey was conducted in 2010 to develop the models for the Dhaka Urban Transport Network Development Study (DHUTS) (JICA, 2010). The third dataset was collected in 2019 to prepare a feasibility study and preliminary design for constructing the Dhaka subway (TYPSA, 2019). The 2005 and 2010 datasets have been processed and analyzed previously by Flavia and Choudhury (2019) to compare the temporal transferability of discrete choice and count regression models in the context of vehicle ownership decisions.

The study area and methodology for conducting the 2005 and 2010 household surveys were the same, albeit the 2005 STP dataset

Table 1
Household Survey Vehicle Ownership Dataset.

Choices	2005 data-STP (World Bank, 2005)		2010 data-DHUTS (JICA, 2010)		2019 data-Subway (TYPSA, 2019)	
	Observations	Percentage (%)	Observations	Percentage (%)	Observations	Percentage (%)
Car	33	5.04	957	5.29	1079	4.63
Car +	3	0.46	115	0.64	120	0.51
Motorcycle	26	3.97	516	2.85	2027	8.70
Bicycle	7	1.07	167	0.92	1061	4.55
No vehicle ownership	586	89.47	16,329	90.3	19,022	81.61
Total	**655**	**100**	**18,084**	**100**	**23,309**	**100**

Table 2
Socio-demographic features in the datasets.

Explanatory Variables	2005 Dataset Average	2010 Dataset Average	2019 Dataset Average
Household size	4.23	4.00	3.94
Number of workers	1.37	1.38	1.32
Number of children	1.23	1.41	1.18
Monthly household income (in 1000 s BDT, 1 BDT = 0.012 USD)	22.7	31.8	37.15

consists of observations from 655 households, whereas the 2010 DHUTS study consists of 18,084 households. On the other hand, the 2019 household survey covered a greater area, i.e. Dhaka Regional Area, which consequently led to a higher sample of 33,178 households. However, in our study, we select a portion of the dataset to conform to the same area where the 2005 and 2010 datasets were collected. For more details on the sampling methodology and the study area of the surveys, readers can refer to JICA (2010); TYPSA (2019); World Bank (2005); and Zannat et al. (2021).

As seen in Table 1, between the years 2005 and 2010, there has been a slight increase in the percentage of households owning a car and more than one car (referred as car+). On the other hand, there has been a slight decrease in the ownership of motorcycles and bicycles between 2005 and 2010. However, the smaller size of the 2005 dataset makes it difficult to make a firm conclusion. It may be noted that no significant changes in the transport infrastructure and public transport provision occurred over the 5-year period of this study in Dhaka to substantially alter vehicle ownership preferences (Flavia and Choudhury, 2019).

However, between the years 2010 and 2019, there has been a rapid increase in the number of motorcycles and bicycles along with a decrease in the number of cars. This has led to a rapid decrease in the total number of households that do not own a vehicle. The possible reasons for this change could be the increased expenses for maintaining a car, change of preferences towards vehicles and increase in the growth of the middle class who are more likely to own motorcycles than cars (Asjad, 2020). More importantly, app-based ride-hailing applications using motorcycles were introduced in the city in 2016, which as reported by Wadud (2020) has led to a substantial increase in the number of motorcycles in the city.

The use of the two prediction datasets thus enables us to compare prediction performances in a moderately different scenario (2005) and a radically different setting (2019) compared to the training dataset (2010).

The explanatory variables available in the dataset consist of the socio-economic characteristics of the household including the size of the household, number of earners/workers, children, and the monthly household income. Table 2 shows the variation of the explanatory variables between the datasets.

In the 2005 and 2010 datasets, monthly household income is collected on a continuous scale, but in the 2019 dataset, income has been collected in 6 income bands. To convert the incomes reported as categorical variables (i.e. income bands) into a continuous scale, the mid-point of each income category has been used. This transformation, however, leads to potential loss in the richness and the quality of data collected in the 2019 dataset.

It is observed that the average household size has slightly decreased between 2005 and 2019. The number of workers remained similar between 2005 and 2010; however, there has been a slight decrease in the year 2019. The average number of children in a household increased between 2005 and 2010; however, it decreased between 2010 and 2019. The average monthly household income increased from 22,700 to 37,200 Bangladeshi Taka (BDT) between the year 2005 and 2019.

4. Model development

As mentioned in Section 3, it can be observed that the 2005 dataset is much smaller compared to the 2010 dataset and some choices are poorly represented (e.g. with only 3,7 observations, see Table 1). On the other hand, income is specified income bands in the 2019 dataset, and the coarse granularity may affect the income coefficients in the model. For the above-mentioned reasons, the 2010 dataset is deemed more suitable for model development, and the 2005 and 2019 datasets are used for external validation. The details of the model development are presented below.

4.1. Choice models (CM) specifications

Previous studies in the Global South context indicate that household income is the most important predictor for vehicle ownership (Choudhary and Vasudevan, 2017; Zegras and Gakenheimer, 2006). Some studies show that larger households and the presence of children would result in more travel needs resulting in a greater need to own vehicles (Choudhary and Vasudevan, 2017; Li et al., 2010). Contrary to this, Zegras and Gakenheimer (2006) posit that a larger household incurs more expenses and fewer savings to purchase a vehicle. More workers in the household are also considered to encourage vehicle ownership as there would be more commuting needs (Flavia and Choudhury 2019).

Based on these findings from the literature, the candidate variables for the vehicle ownership model in the Global South context include household income, household size and composition. Among these, income is expected to have the strongest and possibly a nonlinear effect on the utility. In other words, the utility derived for unit income by a low-income household is not expected to be the same compared to the utility derived by a high-income household (Lerman and Ben-Akiva, 1976). This prompted us to test different specifications with non-linear effects in our research, which included specifying income on a logarithm scale, on a log-linear scale (gamma transform), in thresholds and using piecewise linear transformations. The thresholds and the piecewise linear transform were set to vary in the three different income groups, i.e. low, middle and high income groups using the bottom 40th percentile, 40th to 90th percentile and the top 10th percentile of household income distributions as the demarcation points, respectively[3].

The presence of children in a household was included as a dummy variable. All of the remaining variables, i.e. household size and number of workers, were added as a continuous metric. Initially, all the β coefficients were considered to be generic for all the choices. However, alternative specific β coefficients yielded improved model specifications. Appendix II further shows the utility functions used in this study. All models were estimated using the choice modelling library Apollo (Hess and Palma, 2019) in the statistical software R.

Table A1 in the appendix shows the summary of the goodness of fit metrics observed for the different models. It can be observed that the piecewise linear transform has the best fit with the lowest AIC, BIC, and maximum adjusted rho square and log-likelihood. Hence, this model is considered the most appropriate.

4.2. Machine learning techniques

As explained earlier, ML techniques are highly sensitive to hyperparameter tuning and different initialization settings. Therefore, the optimal hyperparameters were selected by evaluating a 10-fold cross-validation, which was repeated ten times under different initialization settings. After obtaining the optimal hyperparameters, the ML model were trained using the entire training dataset and evaluated a hundred times on different initialization settings. Hence, the goodness of fit metrics and market shares are stated with mean scores and standard deviations. Partial dependence plots were also obtained by averaging the responses of a hundred training sets. All the ML models were computed on the Sci-kit Learn library in Python (Pedregosa et al., 2011).

4.2.1. Neural networks

The hyperparameter grid evaluated for Neural Networks can be observed in Table A2. The activation function was fixed to Rectified Linear Unit (ReLU), and the optimizer for stochastic gradient descent was set as Adaptive Moment Estimation (ADAM), both of which are customary in ML applications for modelling discrete outcomes (Wang et al., 2020a; Han et al., 2022). The number of iterations was set to be 125 iterations based on trial and error. Only L2 regularization has been used as it is the most commonly used regularization technique (Goodfellow et al., 2017). Since the input layer consists of only five variables, and to avoid overfitting, it was deemed appropriate that the number of hidden layers does not exceed 3 and the number of neurons in each hidden layer varies between 5 and 20. The dataset was also pre-processed by standardizing for better convergence and optimization in the estimation of the neural network's weight and biases (Hastie et al., 2009).

4.2.2. Gradient boosting trees

Gradient Boosting Trees is regularized using two interlinked features, i.e. the number of trees and the learning rate (Hastie et al., 2009). The number of trees indicates the number of iterations to be performed, controlled by early stopping. If the loss score, i.e. log-likelihood, on a validation set, does not change by a set tolerance for a certain number of iterations, the training process stops. In this study, early stopping was used; the learning rate and another important hyperparameter, i.e., the depth of the tree, was found using hyperparameter grid search. It is also recommended that after training of one tree, a sample of the total training dataset should be used for the next training. This hyperparameter, called a subsample, was set to 0.8, which is usually recommended for small datasets (Hastie et al., 2009). The hyperparameter grid evaluated for GBT in this study is also presented in the appendix Table A2.

The metrics used to evaluate the ML techniques and select the hyperparameters are usually accuracy, precision or recall. However, as metrics based on prediction are of limited interest to choice modellers, this study uses log-likelihood and mean absolute percentage error (MAPE) of market shares (MS). MAPE of MS is considered a valid metric as it complements LL (LL is already being minimized as the loss term in both NN and GBT) and considers how well the model retained the aggregate market shares. Other probabilistic metrics

[3] These were the most appropriate distribution of income groups in Dhaka according to a study by Rahman (2016) and performed better than the thresholds used by Flavia and Choudhury (2019) and JICA (2010). It can be argued that income could have been divided into more groups, however, that would lead to a loss in the degree of freedom. It is also not possible to check all the possible ranges and the threshold values so there is a need to set the thresholds arbitrarily (Ben-Akiva and Lerman, 1985).

Table 3
Cross-Validation score of ML models selected on different metrics.

Performance Metric	Optimal Hyperparameters based on LL	Optimal Hyperparameters based on MAPE of MS
Neural Networks	Depth = 3, Width = 20, L2 reg = 0.001	Depth = 1, Width = 10, L2 reg = 0.001*
Average Log-likelihood	−0.3638	−0.364
MAPE of MS (%)	6.91	3.48
Gradient Boosting trees	Depth = 2, Learning Rate = 0.05	Depth = 1, Learning Rate = 0.05*
Average Log-likelihood	−0.3638	−0.3641
MAPE of MS (%)	−3.79	−2.89

* Selected hyperparameters.

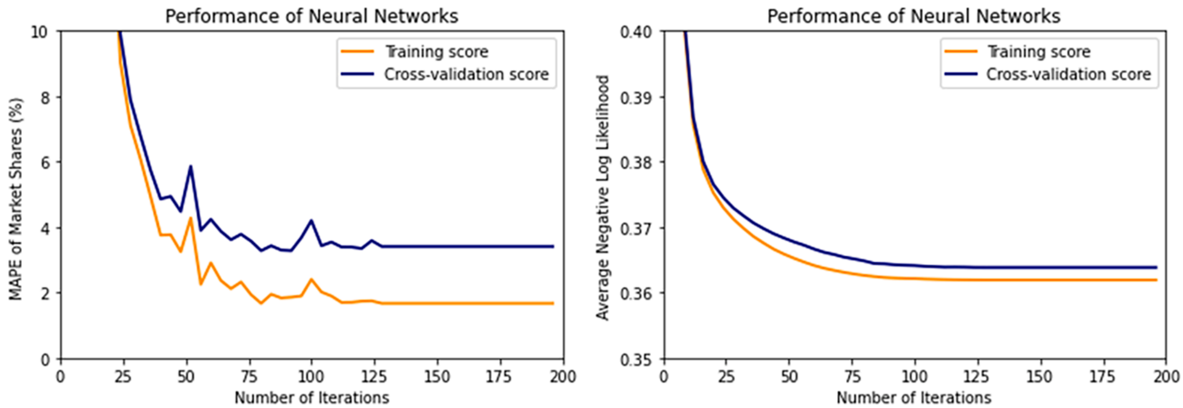

Fig. 2. The learning curve of selected Neural Network.

Table 4
Comparison of the goodness of fit metrics of CM and ML models.

Dataset	Goodness of fit metric	MNL Piecewise Linear Transform	Neural Networks		GBT	
			Mean	Std. Dev	Mean	Std. Dev
2010 Training Dataset (n = 14467)	MAPE of MS (%)	0	2.22	0.94	1.00	0.32
	Log-likelihood	−5251.6	−5239.8	17.6	−5216.19	11.56
2010 Testing Dataset (n = 3617)	MAPE of MS (%)	**0.78**	2.39	0.94	0.92	0.35
	Log-likelihood	**−1312.3**	−1314.51	3.67	−1326.98	2.36
2005 Datasets (n = 655)	Log-likelihood	−265.94	**−253.32**	1.17	−262.41	0.96
	MAPE of MS (%)	30.83	**10.23**	1.33	25.81	0.71
2019 Dataset (n = 23309)	Log-likelihood	**−16555.0**	−17237.98	194.24	−16597.19	52.49
	MAPE of MS (%)	**39.85**	58.12	5.96	40.21	1.76

used in the literature (Hillel, 2019) have been the average probability of correct assignment (APCA) and the negative exponent of LL. APCA was not considered as it is dependent upon predictions. Similarly, the negative exponent of LL was also not considered as it converts the LL values as positive and increases the difference between two similar LL scores.

The optimal hyperparameters and their performance on the cross-validation dataset are presented in Table 3. As expected, the performance of the ML model selected on LL has a higher LL in the cross-validation scores, whereas the ML model selected based on MAPE of MS has a lower MAPE in the cross-validation score. However, the difference between the LL scores is not very high compared to the difference in the MAPE of MS. For example, for the NN, there would be a difference of 1.81 in the log-likelihood score in the testing dataset, which consists of 3617 observations; however, there would be a difference of approximately 3.5 % in the MAPE of MS. Therefore, the ML models' hyperparameters are selected based on the lowest MAPE of market shares. The optimum neural network based on MAPE of MS was found to be one hidden layer constituting ten neurons and a regularization of 0.001. Similarly, the optimum depth of the GBT was found to be one, which is essentially a stump or a single leaf with two nodes. This indicates that there is limited interaction between the different explanatory variables in the dataset.

Fig. 2 shows the relationship between the number of iterations with the metrics average negative LL and MAPE of MS for NN. The figure shows that LL becomes stable at nearly 100 iterations though the MAPE of MS becomes stable after 125 iterations; therefore, setting the number of iterations as 125 is appropriate. The flattening of both the MAPE of MS and the LL indicates that the probability estimates of the ML model have become stable.

Table 5
Comparison of market segments obtained in CM and ML models in the testing dataset.

Market Segment	Choice	Actual Market Shares (%)	MNL Piecewise Linear (%)	Neural Network (%)		Gradient Boosting Trees (%)	
				Mean	Std Dev	Mean	Std Dev
Low Income Group	Car	0.68	0.72	0.64	0.09	0.64	0.09
	Car +	0.07	0.04	0.04	0.01	0.04	0.01
	Motorcycle	0.96	0.98	1.31	0.17	1.30	0.17
	Bicycle	1.23	0.99	0.93	0.07	0.93	0.07
	No Vehicle	97.06	97.27	97.08	0.22	97.09	0.22
Middle Income Group	Car	5.01	5.08	5.10	0.19	5.10	0.20
	Car +	0.46	0.39	0.35	0.03	0.35	0.03
	Motorcycle	3.64	3.92	3.68	0.20	3.68	0.21
	Bicycle	0.68	0.85	0.93	0.08	0.92	0.08
	No Vehicle	90.22	89.76	89.94	0.33	89.95	0.35
High Income Group	Car	23.62	23.62	23.41	0.63	23.39	0.64
	Car +	3.52	3.83	4.04	0.21	4.04	0.21
	Motorcycle	6.28	4.91	5.10	0.30	5.07	0.30
	Bicycle	0.75	0.99	1.12	0.16	1.11	0.15
	No Vehicle	65.83	66.65	66.34	0.80	66.37	0.80
MAPE (%)		–	**12.44**	**18.31**	**1.89**	**31.02**	**2.21**
Weighted MAPE (%)		–	**1.03**	**1.27**	**0.26**	**1.44**	**0.16**

5. Comparison of the CM and ML models

In this section, a comparison of CM and ML models is provided based on the methodology described in Section 2.5.

5.1. Prediction performance

The comparison of the goodness of fit metrics of the CM and ML models is summarized in Table 4. The model with the best LL and MAPE are presented in bold-face.

5.1.1. Testing

On the testing dataset, the performance of MNL is found to be the best, followed closely by NN with a difference of 2.1 in the log-likelihood scores, which indicates that the ML techniques have not achieved a better fit than the CM in this dataset. This strengthens the findings of Nam et al. (2017), Hillel (2019) and Wang et al. (2020b), where the performance of the ML techniques, in terms of LL, is at par or lower than the CM. The NN and GBT models are also able to reasonably predict the overall market shares in the testing dataset with a MAPE of 2.37 % and 0.92 %, respectively. However, MNL again has the lowest MAPE.

Table 5 presents the predicted market shares in the different income groups in the testing dataset. It is observed that the NN model is able to capture vehicle ownership in the different market segments with an overall MAPE of 18.31 %. Still, the performance of the GBT is poor, with a MAPE of 31.02 %.

The other metric used to compare the market segments is the weighted MAPE, where each choice's MAPE is weighted by the number of observations present for that particular choice in the dataset. It is observed that MNL has the lowest weighted MAPE of 1.03 %, followed closely by NN of 1.27 % and 1.44 % for GBT. This indicates that the GBT does not perform well within the market segments compared to the NN. This finding is contrary to the results of Zhao et al. (2019) and Hillel (2019), where the better performing ML algorithms in predicting market segments was GBT compared to NN. The reason why GBT model has not been able to retain the market segments and has a lower log-likelihood is potentially due to overfitting the training dataset.

Nevertheless, it must be highlighted that while building the MNL model, the different income-based market segments have been directly incorporated in the utility function. In contrast, the ML models have 'learned' it automatically. This finding confirms that ML techniques can reduce the risk of model misspecification as they are not dependent upon the modeller's intuition and specification.

5.1.2. Predictions using external datasets

For the backcasting prediction scenario (using the 2005 dataset), the MNL model is found to have the worst MAPE for market share (30.83 %) compared to NN (10.23 %) and GBT (25.81 %). The LL values also follow the same trend. The performance of NN is hence found to be better compared to MNL and GBT in this scenario, which indicates that the NN model is the most temporally transferable model for the 2005 dataset. One of the possible reasons for better prediction of the NN model is that there are no substantial differences between the 2005 and 2010 datasets. It may be noted that the findings regarding the MNL models are similar to the results found in Flavia and Choudhury (2019), where it was found that though some individual taste parameters of the CM of vehicle ownership were transferable between 2005 and 2010, overall, the models were not temporally transferrable.

For the forecasting prediction scenario (using the 2019 dataset), all models have a worse prediction performance of market shares (in comparison with 2005). This is not unexpected given the substantial changes in vehicle ownership between 2010 and 2019. However, comparison of the relative performance of the three models reveals a reverse trend from the backcasting scenario, with the

Table 6
Estimation results of MNL piecewise linear transform model for the 2010 dataset.

Explanatory Variables	Car		Car +		Motorcycle		Bicycle	
	Estimate	Rob t-ratio	Estimate	Rob t-ratio	Estimate	Rob t-ratio	Estimate	Rob t-ratio
$ASC_{\text{Owning vehicle}}$	−5.930	−11.52	−9.167	−5.98	−7.326	−14.29	−5.440	−10.86
$\beta income_{low}$	0.128	4.25	0.052	0.60*	**0.203**	**7.03**	−0.023	−1.03*
$\beta income_{middle}$	0.064	24.44	**0.091**	**8.43**	0.017	4.59	−0.001	−0.20*
$\beta income_{high}$	0.003	3.17	**0.005**	**5.31**	−0.001	−0.18*	**0.006**	**3.03**
$\beta workers$	−0.010	−0.19*	0.305	2.82	0.015	0.24*	0.380	3.35
$\beta householdsize$	−0.173	−4.24	−0.036	−0.36*	0.046	0.87*	0.027	0.26*
$\beta children$	0.258	2.45	0.536	1.70	0.085	0.65*	0.644	2.55
Number of Observations		14,467						
LL (initial)		−23283.74						
LL (final)		−5251.60						
Adj. Rho Square		0.773						

*Statistically insignificant at 90 % confidence interval.

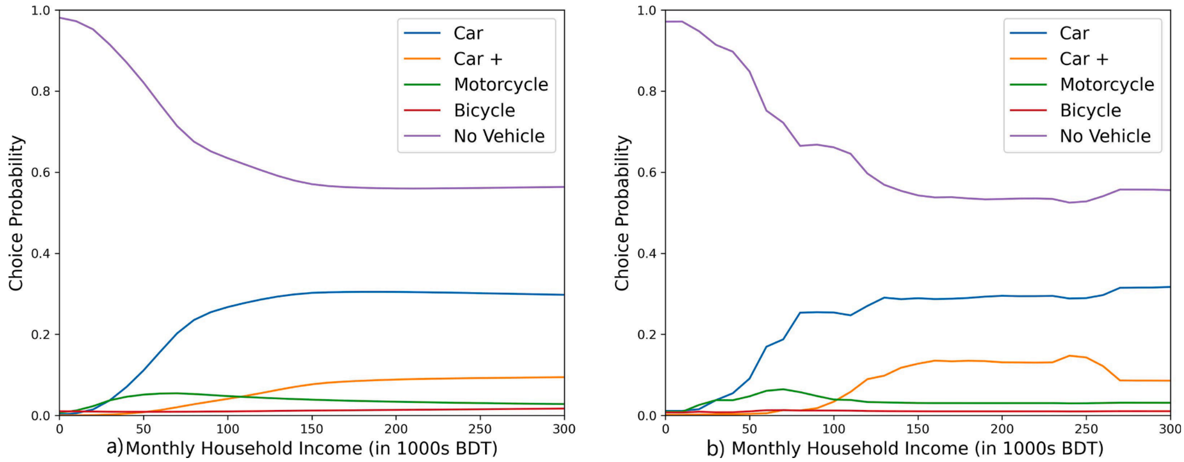

Fig. 3. Partial Dependence Plot for (a) Neural Networks and (b) GBT with respect to monthly household income [1 BDT = 0.012 USD].

MNL model having the best MAPE for market share (39.85 %) followed by GBT (40.21 %) and NN (58.12 %).

On closer look (Table A3 in the Appendix), it is observed that all models underestimate the presence of motorcycles and bicycles and over predict the total number of cars. This is not surprising - as mentioned in Section 3, the introduction of the app-based motorcycle ride-hailing services, there was a substantial increase in the number of motorcycles and reduction in car-ownership (see Wadud, 2020 for details).

5.2. Interpretation

The interpretation of the model results is divided into two sections: household income and other explanatory attributes. Household income has been explained in detail as it exhibits a nonlinear response and is the most important explanatory variable in vehicle ownership decisions.

5.2.1. Household income

Coefficients of the CM (presented in Table 6) indicate that all else being equal, there is a propensity not to own a vehicle. There is significant heterogeneity in vehicle ownership depending on income. Among the low-income people (income less than 18,000 BDT per month), the propensity to own a motorcycle is the highest. A potential reason may be the fact that motorcycles are a common mode for delivery of couriers and most people working in this sector are from the low-income group. The medium income people (income 18,000 BDT to 65,000 BDT) have a higher propensity to own one car, while the high-income people (income higher than 65,000 BDT) have a higher propensity to own more than one car and bicycles. It may be noted that bicycles are frequently used as a recreational vehicle as opposed to main modes – hence it is not unusual that high-income people have a higher propensity to own bicycles.

For ML models, partial dependence plots with respect to monthly household income for NN and GBT are presented in Fig. 3. The horizontal scale has been set to a monthly household income up to 300,000 BDT to show the dependence of income at outliers. Partial dependence plots of both NN and GBT exhibit an intuitive substitution pattern of alternatives, the nonlinear effect of income and a saturation level for the choice probabilities.

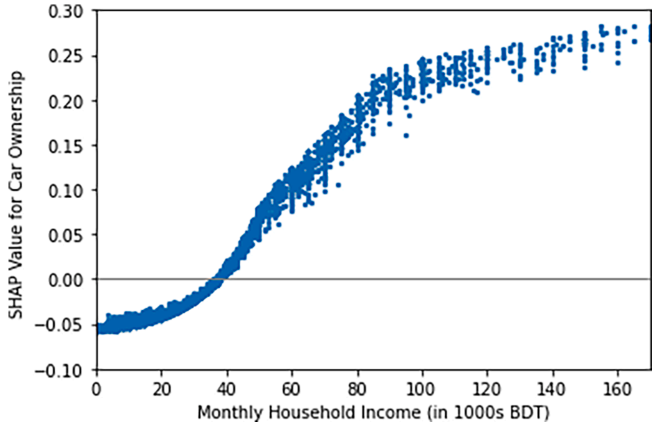

Fig. 4. SHAP Value plot of car ownership and monthly household income in Neural Networks for the year 2010.

Table 7
Comparison of Income Elasticities in CM and ML models.

Elasticity	MNL Piecewise Linear transform	Neural Network		Gradient Boosting Trees	
		Mean	Std Dev	Mean	Std Dev
Car	1.287	1.397	0.042	1.103	0.089
Car +	1.549	1.839	0.058	1.620	0.239
Motorcycle	0.594	0.556	0.040	0.616	0.078
Bicycle	−0.125	0.004	0.070	0.280	0.102
No Vehicle	−0.111	−0.120	0.004	−0.104	0.007

The choice probability of a motorcycle follows an inverted U style curve, where the probability increases up to a monthly income of 60,000 BDT, after which it starts to decrease. This is intuitive as it is observed in developing countries that households purchase cars in favour of motorcycles once they have the financial ability to do so due to the extra convenience offered by cars in long distance/time trips and the poor safety record of motorcycles (Gwilliam, 2003; Law et al., 2015) This also highlighted in MNL model in Table 6, where the coefficient for owning a motorcycle for the high-income group is negative and statistically insignificant at 90 % confidence interval.

The probability of owning more than one car is intuitive as the choice probability starts to increase after a monthly household income of 60,000 BDT and reaches a saturation level. The MNL model results, presented in Table 6**,** show that income for the lower-income groups for owning more than one car is statistically insignificant. However, the coefficients are positive and statistically significant for the middle and high-income groups.

The probability of owning a bicycle largely remains stable though it can be observed in Fig. 3a that at a high monthly household income of 400,000 BDT, the probability of choosing bicycles steadily rises and matches up to the probability of owning a motorcycle. A similar result is also obtained in the MNL model, where surprisingly, the high-income group has a higher preference for bicycles.

The MNL model also estimates the marginal effect of income on car ownership as the coefficients reduce from low-income to high-income groups. In Fig. 3, it can be observed that there is a rapid increase in the probability of car ownership between the range of 40,000 BDT and 80,000 BDT, after which the rate of increase of probability starts to reduce and ultimately reaches a saturation level. The car ownership curve also matches the theoretical S-shaped car ownership curve indicated by Zegras and Gakenheimer (2006) and supports that there is an existence of an income threshold for car ownership.

Fig. 4 shows the effect of monthly household income on the probability of owning a car using SHAP values for the NN model. It can be observed that a monthly household income below 40,000 BDT decreases the probability of owning a car since the SHAP values are negative. However, after this threshold, household income starts to have a positive effect on owning a car. Hence, the car ownership threshold is 40,000 BDT for the year 2010. Correlating this finding with the monthly income distribution in the dataset revealed that 75 % of the households were below the car ownership income threshold of 40,000 BDT in the year 2010. This affirms Zegras and Gakenheimer (2006) findings where it is highlighted that most households remain below the car ownership threshold in developing countries. The computation of car ownership threshold and the car ownership probability curve is of utmost importance to forecast car ownership growth in developing countries as forecasts that do not consider the income distribution perform poorly (Gakenheimer, 1999; Storchmann, 2005). However, the car ownership threshold estimated using SHAP values should be further validated by comparing the results obtained from other mentioned methods in Gomez and Cevedo (2013) and Storchmann (2005).

Income elasticities of vehicle ownership calculated from CM and ML techniques have been compared in Table 7. All of the income elasticities are intuitive as for developing countries, it is expected that the car income elasticity should be greater than 1, which indicates rapid motorization with an increase in income (Gakenheimer, 1999). Since all the elasticities are in a similar range and there are no elasticity values to be used as the 'ground truth', it is difficult to reject any elasticity estimates.

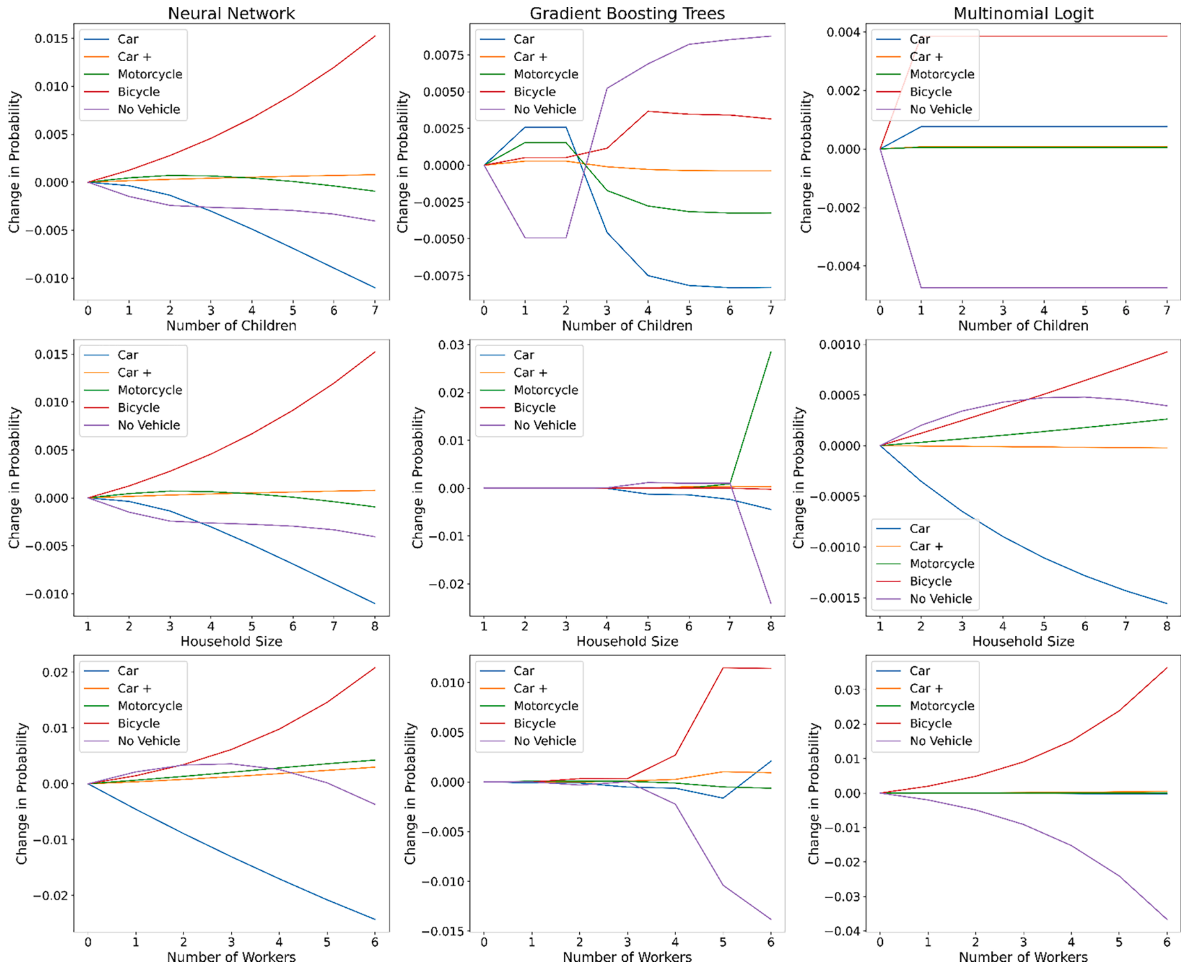

Fig. 5. Change in probability observed in NN, GBT and MNL with a) number of children, b) household size, and c) workers.

5.2.2. *Other explanatory variables*

Partial dependence plots based on the MNL, NN and GBT have also been plotted for the explanatory variables in Fig. 5. The plots have been centred so that only a change in probability is observed rather than the substitution pattern of alternatives.

Overall, it can be asserted that the effect of attributes on the choice probabilities is quite similar in the CM and the ML models in terms of the signs. However, the scales, i.e. the change in probabilities observed, are significantly different. There are also several occasions where the effect of attributes is contradictory to the MNL model results. For example, in Fig. 5c the number of workers has a negative effect on the probability of owning a car in NN, whereas the MNL model shows the opposite.

There are cases where ML techniques can provide further insights as well. For instance, in Fig. 5a, the dependence plot of children in GBT shows that if a household has less than two children, the probability of owning a motorcycle and a car is higher. This is reasonable considering that households with fewer children have more savings. However, when the number of children in a household exceeds two, the probability of owning no vehicle and bicycles increases. This insight could be used to improve the choice models, using a piecewise linear specification or categories for the corresponding variable for instance.

Elasticity estimates were also compared for these explanatory variables, which have been compiled in Table A4 in the appendix. It is observed that in most cases, the elasticity estimates of the ML models differ significantly in both signs and magnitude compared to the CM. However, the closeness of elasticity estimates obtained from NN for income indicates that the ML techniques can indeed provide reliable estimates for several significant variables. Overall the comparison of the elasticity estimates affirms the findings of Zhao et al. (2020) and Wang et al. (2020a), where the scale of elasticity estimates of the ML techniques was significantly different from the CM.

To conclude this section, it can be observed that choice models with their link to economic theories remain invaluable in explaining and interpreting attributes. However, the nonlinear effect of attributes can be better understood and visualized using explainable ML techniques such as partial dependence plots. ML techniques can therefore be used as an exploratory tool to improve the choice model specifications (Basu, 2019; Hruschka et al., 2002; Zhao et al., 2020).

6. Conclusion

Vehicle ownership, especially in the developing countries, is an issue of global interest due to the growing energy and environmental concerns. Better insights about the factors that are affecting the vehicle growth and understanding the underlying heterogeneity in preferences among the individuals is of immense practical importance. In particular, these are critical for designing effective policies that can mitigate the growth of private cars and promote sustainable transportation systems. From a planning and policy making perspective, the robustness of the modelling methodology used for vehicle ownership is crucial. In the current era, where there is a growing interest in machine learning-based modelling approaches, it is important to understand when traditional choice modelling approaches may be more advantageous and when ML models may offer benefits.

This study compares choice models, i.e. MNL and machine learning techniques, i.e. NN and GBT, in the context of vehicle ownership decisions in Dhaka, Bangladesh. While some of the particular findings of the study (e.g. the probability of owning cars is the highest for households with monthly income between 40,000–60,000 BDT and households with less than two children, the income elasticities for owning car and motorcycle are 1.28 and 0.594 respectively) are directly relevant for policy makers in Bangladesh, the wider methodological insights gained from the study is of our principal focus. The key such findings are listed below:

- Results indicate that CM outperforms ML techniques on the testing dataset in terms of the log-likelihood and mean absolute percentage error of market shares. This indicates that ML models do not necessarily have a better goodness of fit compared to the CM. These results are similar to Nam et al. (2017), Hillel (2019) and Wang et al. (2020b), who compare mode choices using similar performance metrics, i.e. log likelihood.
- The MNL and ML models were also used to predict vehicle ownership in different market segments based on income groups in the testing dataset. NN was able to capture the different market segments and hence more suitable to be used for forecasting purposes compared to GBT. However, the best performing model, in terms of lower prediction error, was found to be MNL when compared to the ML models.
- Backcasting and forecasting the market shares (for the years 2005 and 2019, respectively) provide insights about the relative performance of the CM and ML models in substantially different application contexts. We find that for backcasting (where the temporal gap was 5 years and the differences in the market shares were not substantial), neural networks had the best performance. However, for the forecasting (where the temporal gap was 9 years and the transport landscape has substantially changed, particularly due to the introduction of app-based motorcycle ride-hailing services), the MNL model performed relatively better. These conclusions are in line with a discussion paper by Van Cranenburgh et al. (2022) who state that models without behavioral underpinning are expected to have poor forecasting abilities in new contexts (such as an increased temporal gap). Therefore, our study provides an empirical case for using choice modelling techniques for long-term forecasting which is very relevant to policy makers and practitioners. Typically, household travel surveys are conducted after 5 or more years in most countries in the global south, therefore assessing which model performs best for forecasting when there is a large temporal gap is often of immense interest for policy makers.
- Elasticity estimates of the NN and GBT models were also similar to those found in MNL for income. However, for other explanatory variables (household size, number of workers and number of children) the effects of attributes and elasticities were opposite in the ML techniques when compared to the MNL model and non-intuitive. This implies that the ML techniques cannot be relied upon to provide consistent explanations which can be used for policy formulation.
- While comparing choice models and ML techniques, we explore the suitability of using explainable machine learning techniques for generating outputs of interest to planners and policy makers. Our findings reveal that the partial dependence plots can be used to generate intuitive car ownership probability curves (i.e. replicate the S-shaped curve reported in the transport literature (Zegras and Gakenheimer, 2006; Dargay et al., 2007)) and SHAP can be used to reliably determine vehicle ownership income thresholds. The latter can be used by policy makers to predict vehicle ownership under different income growth scenarios.

While it is difficult to fully generalize the conclusions for all other contexts, the fact that our findings are in line with the previous empirical findings and general speculations (outlined in Van Cranenburgh et al. (2022)), it suggests wider applicability of the findings beyond the specific context and dataset. It may be noted though that we compare CM and ML techniques in a choice context where there are limited explanatory variables and the sample size of the training dataset is moderate (14,467 observations). We are therefore cautious in blanket generalization of the findings, especially to contexts where the number of explanatory variables and/or observations are much larger. Also, this research has focused on the comparison of traditional CM and ML models using datasets from a megacity in a developing country. It will be useful to conduct similar studies in the context of cities where the transport landscape has not undergone such radical changes to get further insights about generalisation of the findings.

Further, in the last few years, there has been a growing interest to develop models that combine the CM and ML. Examples include Learning Multinomial Logit (Sifringer et al.,2020), TasteNet-MNL (Han et al., 2022) and Embeddings Multinomial Logit (Arkoudi et al., 2021) – which are hybrids of MNL and NN, ASU-DNN model (Wang et al., 2020a)– a priori behavioural constraint driven NN, etc. In future research, it will be interesting to compare the performance of these new genre of models with the conventional MLs and CMs.

CRediT authorship contribution statement

Azam Ali: Conceptualization, Methodology, Formal analysis, Software, Visualization, Writing – original draft. **Arash Kalatian:** Writing – review & editing. **Charisma F. Choudhury:** Supervision, Conceptualization, Funding acquisition, Project administration, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by a UKRI Future Leaders Fellowship Next Generation Travel Behaviour Models' [Grant Number MR/T020423/1], the Commonwealth Scholarship Commission, United Kingdom and ESRC AQM Grant Reference No. ES/P000746/1.

The authors would like to thank Dr Shenhao Wang for his feedback on the initial stage of this research.

Appendix I. List of studies comparing choice models and machine learning techniques in travel behavior modelling

Reference	Aims	Data	Comparison Methodology/Metrics	Findings
Paredes et al. (2017)	Comparison of ML classifiers (Decision Tree (DT), Random Forest (RF), Gradient Boosting Tree(GBT), Ensemble, Support Vector Machine (SVM)) and Multinomial Logit (MNL) model in terms of prediction accuracy	Household survey data of 2008 and 2012	Build MNL and ML models on 2008 data and predict on 2012 data. Comparison using prediction accuracy	ML models 10 % more accurate in predictions compared to MNL. Hyperparameter tuning required for better ML results
Lee et al. (2018)	Comparison of four types of NN with MNL in terms of prediction accuracy	Mode Choice Data Removes transit and bicycle mode data to remove class imbalance problem	Prediction accuracy Sensitivity analysis, similar to Partial Dependence Plot (PDP)	ML models 10 % more accurate in predictions compared to MNL
Lee et al. (2019)	Use of GBT to explain the uptake of autonomous vehicles	SP data	PDP and Feature Importance Plots used for interpretation	Interpretable ML techniques can reveal intuitive behaviour
Wang and Ross (2018)	Comparison of MNL and GBT performance	Mode choice data Removes transit and bicycle mode data to remove class imbalance problem	Mode specific prediction accuracy	GBT performs better in prediction accuracy on both the training and testing dataset Both MNL and ML techniques don't perform well with imbalanced data in terms of prediction
Zhao et al. (2020)	Comparison of MNL, MMNL and ML techniques (NB, DT, RF, GBT, Bag, SVM, NN)	SP mode choice data	Individual predictions Market Shares (MS) calculated by probabilities Partial Dependence Plots Feature Importance Plots Elasticities, Marginal Effects	Each alternative's (i.e. modes) prediction accuracy is dependent on the number of observations for the alternative indicating aggregation bias RF outperforms MNL with better prediction accuracy and the least deviances in market shares MMNL fits the training dataset better in terms of LL and adj. rho compared to MNL; however, the performance of MNL is better than MMNL on the test data indicating overfitting of the MMNL model ML models inferior in interpretation due to lack of behavioural underpinning
Zhao et al. (2019)	Can ML techniques (Naïve Bayes (NB), DT, RF, GBT, Bagging, Logistic Regression (LR), NN) capture taste heterogeneity in mode switching behaviour?	SP survey	PDP, Individual Conditional Expectation (ICE) Conditional PDP & ICE for separate market segments Elasticities, Marginal	ML techniques can capture heterogeneity which can be visualised using ICE and good fit on different market segments

(*continued on next page*)

(*continued*)

Reference	Aims	Data	Comparison Methodology/Metrics	Findings
			Effects for different market segments	
Hagenauer and Helbich (2017)	Comparison of MNL and six ML algorithm (SVM, NN, Boosting, Bagging, RF, NB)	Mode choice data Resampling carried out to remove class imbalance	Prediction accuracy of each mode Feature Importance Plots	RF performance is best with MNL performing poorest in terms of prediction accuracy
Nam et al. (2017)	Comparison of NL, Cross Nested Logit (CNL) and NN (with different number of layers and activation functions)	Mode choice data (swiss metro)	MS calculated based on probabilities and predictions Log Likelihood (LL)	NN model performed best in terms of lowest RMSE in MS LL of CNL similar to the best performing NN
Lhéritier et al., (2019)	Comparison of MNL, latent class MNL, and RF	Airline itinerary data (highly complex and non-linear)	Top N metric (ranks the probability of the alternatives) Computation time Feature Importance Plots	RF captures nonlinear behaviour, better in computation time and prediction accuracy compared to MNL models
Xie et al., (2003)	Comparison of MNL with DT and NN	Commuting mode choice data Uses weights to cater for imbalanced dataset	Prediction accuracy MS calculated using predictions	NN and DT are better in prediction accuracy compared to MNL
Mohammadian and Miller, (2002)	Comparison of NL and NN	Vehicle type	Prediction accuracy MS calculated using predictions	NN better at prediction accuracy and market shares
Alwosheel et al. (2019)	Use prototypical example to build trust in NN model Compares MNL and NN	Mode choice Handles class imbalance by removing low mode share of bicycles	Prediction accuracy LL Heat map for prototypical examples	Heat map and prototypical examples indicate that NN learns intuitively and in line with a priori beliefs NN performs better than MNL in both LL and prediction accuracy
Cheng et al. (2019)	Comparison of MNL and RF, Adaboost, SVM	Mode choice data	Prediction accuracy MAPE of aggregate MS using predictions Computation time	RF and SVM better in prediction accuracy compared to MNL Computation time similar for RF, Adaboost and MNL, SVM slower
De Carvalho et al. (1998)	Comparison of MNL and NN	Synthetic data And real mode choice data	Root mean square error of prediction/ probabilities (unclear)	NN achieves good fit (lower RMSE) on synthetic data which is nonlinear
Golshani et al. (2018)	Comparison of MNL copula-based model with NN (two types of NN i.e. one for mode choice, another for departure time)	Mode choice and departure time data	Prediction accuracy Sensitivity analysis (similar to PDP)	NN captures nonlinearity better than MNL models but does not allow extraction of policy insights Prediction accuracy better for NN compared to MNL
Omrani, (2015)	Comparison of MNL and NN (simple neural network, radial based function), SVM	Mode choice data	Average probability of correct assignment (APCA)	NN better in APCA than MNL
Bentz and Merunka (2000)	Comparison of MNL and NN Enhance specification of MNL using output of NN	Brand choice decisions	Rho (McFadden R^2) Market shares through predictions	NN better at fitting the data compared to MNL with better rho Nonlinearities are better modelled by NN Enhancing MNL does not result in a large difference in rho
Basu, (2019)	Comparison of MNL and NN Enhance specification of MNL using interpretable ML techniques	Car ownership Caters for class imbalance by resampling	Market shares based on Predictions, (use of accuracy, recall, precision) Computation time Feature Importance Plot	ML better in forecasting compared to CM Enhanced CM at par with ML techniques in terms of prediction accuracy and RMSE of market shares

(*continued on next page*)

(*continued*)

Reference	Aims	Data	Comparison Methodology/Metrics	Findings
Hillel (2019, chap. 7)	Comparison of NL (best out of MNL, CNL) with ML techniques (NN, RF, GBT, SVM, extremely randomized trees) Assisted approach of CM specification using EL	Mode choice data	Accumulated Local Effects which is a deviant of PDP Average probability of correct assignment (APCA) Accuracy LL Market Shares calculated using predictions and draws from probabilities	GBT performs best in accuracy and APCA, LL, market shares calculated using both predictions and draws from probabilities compared to NL Relative difference in LL and prediction accuracy between the different models is relatively low Assisted CM performs better than simple CM
Wang et al. (2020b)	Extraction of economic information from NN and comparison with MNL	Mode choice Singapore and London	Prediction accuracy Elasticity Substitution pattern of alternatives (similar to PDP) Social Welfare Value of time VOT	NN better in prediction accuracy compared to MNL Reliable and richer information in the substitution pattern of alternative of NN compared to MNL models Elasticity estimates from NN are overall unreliable Negative VOT observed for datasets with a smaller number of observations
Wang et al. (2020c)	Use of multitask NN to combine RP and SP study Comparison with NL (constrained and unconstrained)	RP and SP	Prediction accuracy LL Elasticity Substitution pattern of alternatives (PDP)	Multitask NN performs better in prediction accuracy compared to NL LL better for NL models Some elasticities of NN are similar to NL, however, some signs are opposite
Yao and Bekhor (2020)	Combination of MNL and ML clustering for route choice models	GPS Data in Tel Aviv	Prediction accuracy Computation time	Random Forest performed the best in terms of prediction error
Lu et al. (2021)	Comparison of MNL, RF, and Policy gradient reinforcement learning (PGRL)	SP	Prediction accuracy Marginal Effects	PGRL performed best in terms of prediction accuracy
Salas et al. (2022)	Comparison of MNL, MMNL, NN, SWM, K Nearest Neighbours, RF, Extreme GBT where there is heterogeneity in preference	Synthetic and RP mode choice	Prediction accuracy Feature importance	Best prediction accuracy for NN

Appendix II. Utility specifications

1. Income specified in a Logarithm Scale

$$V_{i,n} = ASC_{i,n} + \beta_{workers_{i,n}} {}^{*} workers + \beta_{hhsize_{i,n}} {}^{*} hhsize + \beta_{children_{i,n}} {}^{*} children + + \beta_{income_{i,n}} {}^{*} \ln(income > 0)$$

for $i = motorcycle, bicycles, car, cars+$

$$V_{bicycle,n} = ASC_{bicycle,n} + \beta_{workers_{bicycle,n}} {}^{*} workers + \beta_{hhsize_{bicycle,n}} {}^{*} hh_{size} + \beta_{children_{bicycle,n}} {}^{*} children(dummy) + \beta_{income,bicycle} {}^{*} \ln(income > 0)$$

$V_{novehicles,n} = 0$ (set as base)

$\beta_{drivers\_0_{i,n}} {}^{*}(driver = 0)$ is set as base

2. Income specified in a Log-Linear/Gamma/Box-Cox Transform

$$V_{i,n} = ..\beta_{income_{i,n}} * \left(\gamma^{*} income + (1 - \gamma)\ln(income > 0)\right)$$

3. Income specified in thresholds

$$V_{i,n} = ..\beta_{inc_{low\,i,n}} * income_{low} + \beta_{inc_{middle_{i,n}}} * income_{middle} + \beta_{inc_{high_{i,n}}} * income_{high}$$

where $\beta_{inc_{low}}$ is set as base

4. Income specified using piecewise linear transformation

$$V_{i,n} = ..\beta_{inc_{low_{i,n}}} * (threshold_1) + \beta_{inc_{middle_{i,n}}} * (threshold_2) + \beta_{inc_{high_{i,n}}} * (threshold_3)$$

where

$$threshold_1 = \left\{ \begin{array}{ll} income & \text{if income} < 18,000\text{ BDT} \\ 18,000BDT & \text{otherwise} \end{array} \right\}$$

$$threshold_2 = \left\{ \begin{array}{ll} 0 & \text{if income} < 18,000\text{BDT} \\ income - 18,000BDT & 18,000BDT \leq income < 65,000BDT \\ 47,000BDT & otherwise \end{array} \right\}$$

$$threshold_3 = \left\{ \begin{array}{ll} 0 & \text{if income} < 65,000\text{ BDT} \\ income - 65,000BDT & \text{otherwise} \end{array} \right\}$$

Table A1
Comparison of goodness of fit of estimated MNL models.

MNL Model Name	Log-Likelihood	No of Parameters	Adj Rho Square	AIC	BIC
Logarithm scale	−5303.943	20	0.7713	10647.89	10799.5
Thresholds	−5418.584	24	0.7662	10885.17	11067.1
Piecewise Linear Transform	−5251.602	28	0.7732	10559.2	10771.4
Gamma Transform = 0.5	−5651.027	20	0.7564	11342.05	11493.7
Gamma Transform = 0.25	−5638.21	20	0.757	11316.42	11,468

Table A2
Hyperparameter Grid for Neural Networks and Gradient Boosting Trees.

Hyperparameter Grid	
Neural Network	
Depth	[1,2,3]
Width	[5,10, 15, 20]
L2 Regularization	[0.0001,0.001,0.001,0.01,0.1]
Gradient Boosting Trees	
Depth	[1,2,3]
Learning Rate	[0.01,0.05,0.1,0.15,0.20,0.25,0.30]

Table A3
Comparison of detailed prediction results.

Dataset	Choice	Original Data	MNL Piecewise Linear	NN		GBT	
				Mean	Std. Dev	Mean	Std. Dev
2005 Datasets (n = 655)	Car	33	16.99	32.25	0.87	19.66	37.75
	Car +	3	1.55	**3.39**	0.17	2.52	12.48
	Motorcycle	26	15.66	**19.8**	0.67	15.02	19.47
	Bicycle	7	6.13	**6.26**	0.24	5.2	20.23
	No vehicle ownership	586	614.68	**593.27**	1.56	612.61	52.36
2019 Dataset (n = 23309)	Car	1079	**1479.43**	1701.31	81.37	1522.25	0.42
	Car +	120	**135.61**	208.88	34.52	105.45	0.08
	Motorcycle	2027	**845.8**	60.287	47.4	804.05	0.22
	Bicycle	1061	188.59	208.54	23.89	**242.22**	0.13
	No vehicle ownership	19,022	20659.6	**20587.4**	94.4	20,635	0.51

Table A4
Comparison of Elasticity estimates of MNL, NN and GBT.

Elasticities	MNL Piecewise Linear	NN		GBT	
		Mean	Std. Dev	Mean	Std. Dev
Household size					
Car	−0.498	−0.083	0.024	−0.083	0.024
Car +	−0.002	−0.051	0.055	−0.050	0.055
Motorcycle	0.106	0.015	0.038	0.016	0.038
Bicycle	0.020	−0.020	0.060	−0.019	0.060
No Vehicle	0.341	0.004	0.001	0.004	0.001
Children					
Car	0.120	−0.025	0.020	−0.047	0.017
Car +	0.025	0.017	0.055	−0.016	0.010
Motorcycle	0.019	0.003	0.033	−0.034	0.017
Bicycle	0.053	0.153	0.057	0.057	0.042
No Vehicle	−0.281	−0.001	0.001	0.003	0.001
Number of Worker					
Car	−0.073	−0.089	0.026	−0.090	0.026
Car +	0.114	0.064	0.074	0.065	0.073
Motorcycle	0.014	0.021	0.044	0.020	0.044
Bicycle	0.218	0.203	0.094	0.208	0.092
No Vehicle	−0.345	0.001	0.001	0.001	0.001

References

Aboutaleb, Y.M., Danaf, M., Xie, Y., Ben-Akiva, M., 2021. Discrete Choice Analysis with Machine Learning Capabilities 1–14. [WWW Document]. URL https://arxiv.org/abs/2101.10261 (accessed 20.4.21).

Alwosheel, A., van Cranenburgh, S., Chorus, C.G., 2019. 'Computer says no' is not enough: using prototypical examples to diagnose artificial neural networks for discrete choice analysis. J. Choice Model. 33, 1–13. https://doi.org/10.1016/j.jocm.2019.100186.

Anowar, S., Eluru, N., Miranda-Moreno, L.F., 2014. Alternative modeling approaches used for examining automobile ownership: a comprehensive review. Transp. Rev. 34, 441–473. https://doi.org/10.1080/01441647.2014.915440.

Arkoudi, I., Azevedo, C.L. and Pereira, F.C., 2021. Combining Discrete Choice Models and Neural Networks through Embeddings: Formulation, Interpretability and Performance. [WWW Document]. arXiv preprint arXiv:2109.12042 (accessed 03.03.23).

Asjad, T., 2020. To have or not to have a private car. [WWW Document]. URL https://thefinancialexpress.com.bd/views/to-have-or-not-to-have-a-private-car-1583506811 (accessed 20.1.22).

Basu, R., 2019. Reinterpreting vehicle ownership in the era of shared and smart mobility. MSc Thesis, Massachusetts Institute of Technology.

Ben-Akiva, M., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand, MIT Press. Cambridge, Massachusetts. doi: 10.4324/9780203891506-4.

Bentz, Y., Merunka, D., 2000. Neural networks and the multinomial logit for brand choice modelling: a hybrid approach. J. Forecast. 19, 177–200. https://doi.org/10.1002/(SICI)1099-131X(200004)19:3<177::AID-FOR738>3.0.CO;2-6.

Bhat, C.R., Pulugurta, V., 1998. A comparison of two alternative behavioral choice mechanisms for household auto ownership decisions. Transp. Res. Part B: Methodol. 32, 61–75. https://doi.org/10.1016/S0191-2615(97)00014-3.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning, Springer Science + Business Media, LLC, Singapore.

Cheng, L., Chen, X., de Vos, J., Lai, X., Witlox, F., 2019. Applying a random forest method approach to model travel mode choice behavior. Travel Behav. Soc. 14, 1–10. https://doi.org/10.1016/j.tbs.2018.09.002.

Choudhary, R., Vasudevan, V., 2017. Study of vehicle ownership for urban and rural households in India. J. Transp. Geogr. 58, 52–58. https://doi.org/10.1016/j.jtrangeo.2016.11.006.

Dargay, J., Gately, D., Sommer, M., 2007. Vehicle Ownership and Income Growth, Worldwide: 1960–2030. Energy J. 28, 143–170.

de Carvalho, M.C.M., Dougherty, M.S., Fowkes, A.S., Wardman, M.R., 1998. Forecasting travel demand: a comparison of logit and artificial neural network methods. J. Oper. Res. Soc. 49, 717–722. https://doi.org/10.1057/palgrave.jors.2600590.

de Jong, G., Fox, J., Daly, A., Pieters, M., Smit, R., 2004. Comparison of car ownership models. Transp. Rev. 24, 379–408. https://doi.org/10.1080/0144164032000138733.

Doshi-Velez, F., Kim, B., 2017. Towards A Rigorous Science of Interpretable Machine Learning [WWW Document]. URL http://arxiv.org/abs/1702.08608 (accessed 20.4.21).

Flavia, A., Choudhury, C., 2019. Temporal transferability of vehicle ownership models in the developing world: case study of Dhaka. Bangladesh. Transp. Res. Rec. 2673, 722–732. https://doi.org/10.1177/0361198119836760.

Friedman, J., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232. https://doi.org/10.2307/2699986.

Gakenheimer, R., 1999. Urban mobility in the developing world. Transp. Res. Part A: Policy Pract. 33, 671–689. https://doi.org/10.1016/S0965-8564(99)00005-1.

Goldstein, A., Kapelner, A., Bleich, J., Pitkin, E., 2015. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. J. Comput. Graph. Stat. 24 (1), 44–65.

Golshani, N., Shabanpour, R., Mahmoudifard, S.M., Derrible, S., Mohammadian, A., 2018. Modeling travel mode and timing decisions: Comparison of artificial neural networks and copula-based joint model. Travel Behav. Soc. 10, 21–32. https://doi.org/10.1016/j.tbs.2017.09.003.

Gomez, J., Cevedo, J., 2013. Studying Car and Motorcycle Ownership Levels in Developing Countries Using Individual Income, in: 13th WCTR, July 15–18, 2013. Rio de Janeiro, Brazil, pp. 1–18.

Goodfellow, I., Bengio, Y., Courville, A., 2017. Deep Learning. MIT Press.

Gwilliam, K., 2003. Urban transport in developing countries. Transp. Rev. 23, 197–216. https://doi.org/10.1080/01441640309893.

Hagenauer, J., Helbich, M., 2017. A comparative study of machine learning classifiers for modeling travel mode choice. Expert Syst. Appl. 78, 273–282. https://doi.org/10.1016/j.eswa.2017.01.057.

Han, Y., Zegras, C., Pereira, F.C., Ben-Akiva, M., 2022. A neural-embedded discrete choice model: learning taste representation with strengthened interpretability. Transp. Res. Part B: Metho. 163, 166–186. https://doi.org/10.1016/j.trb.2022.07.001.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning data mining, inference, and prediction. Encycl. Syst. Biol. 508. https://doi.org/10.1007/978-1-4419-9863-7_941.

Hensher, D.A., Rose, J., Greene, W.H., 2005. Applied Choice Analysis: A Primer. Cambridge University Press, New York. doi: 10.1017/CBO9781107415324.004.

Hess, S., Palma, D., 2019. Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. J. Choice Model. 32, 1–20. https://doi.org/10.1016/j.jocm.2019.100170.
Hillel, T., 2019. Understanding Travel Mode Choice: A New Approach for City Scale Simulation. University of Cambridge. Doctor of Philosophy.
Hillel, T., Elshafie, M.Z.E.B., Jin, Y., 2018. Recreating passenger mode choice-sets for transport simulation: a case study of London, UK. Proc. Inst. Civ. Eng. - Smart Infrastruct. Constr. 171, 29–42. https://doi.org/10.1680/jsmic.17.00018.
Hillel, T., Bierlaire, M., Elshafie, M., Jin, Y., 2020. A systematic review of machine learning classification methodologies for modelling passenger mode choice. J. Choice Model. https://doi.org/10.1016/j.jocm.2020.100221.
Hruschka, H., Fettes, W., Probst, M., Mies, C., 2002. A flexible brand choice model based on neural net methodology: Comparison to the linear utility multinomial logit model and its latent class extension. OR Spectr. 24, 127–143. https://doi.org/10.1007/s00291-002-0095-1.
JICA, 2010. DHUTS Dhaka Urban Transport Network Development Study.
Kalatian, A., Farooq, B., 2021. Decoding pedestrian and automated vehicle interactions using immersive virtual reality and interpretable deep learning. Transp. Res. Part C: Emerg. Technol. 124, 102962 https://doi.org/10.1016/j.trc.2020.102962.
Law, T.H., Hamid, H., Goh, C.N., 2015. The motorcycle to passenger car ownership ratio and economic growth: a cross-country analysis. J. Transp. Geogr. 46, 122–128. https://doi.org/10.1016/j.jtrangeo.2015.06.007.
Lee, D., Derrible, S., Pereira, F.C., 2018. Comparison of four types of artificial neural network and a multinomial logit model for travel mode choice modeling. Transp. Res. Rec. 2672, 101–112. https://doi.org/10.1177/0361198118796971.
Lee, D., Mulrow, J., Haboucha, C.J., Derrible, S., Shiftan, Y., 2019. Attitudes on autonomous vehicle adoption using interpretable gradient boosting machine. Transp. Res. Rec. 2673, 865–878. https://doi.org/10.1177/0361198119857953.
Lerman, S.R., Ben-Akiva, M., 1976. Disaggregate behavioral model of automobile ownership. Transp. Res. Rec. 34–51.
Lhéritier, A., Bocamazo, M., Delahaye, T., Acuna-Agost, R., 2019. Airline itinerary choice modeling using machine learning. J. Choice Model. 31, 198–209. https://doi.org/10.1016/j.jocm.2018.02.002.
Li, J., Walker, J.L., Srinivasan, S., Anderson, W.P., 2010. Modeling private car ownership in China: investigation of urban form impact across megacities. Transp. Res. Rec. 76–84. https://doi.org/10.3141/2193-10.
Lu, J., Meng, Y., Timmermans, H., Zhang, A., 2021. Modeling hesitancy in airport choice: a comparison of discrete choice and machine learning methods. Transp. Res. Part A: Policy Pract. 147, 230–250. https://doi.org/10.1016/j.tra.2021.03.006.
Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Nips 16, 426–430.
Mohammadian, A., Miller, E.J., 2002. Nested logit models and artificial neural networks for predicting household automobile choices: comparison of performance. Transp. Res. Rec. 92–100. https://doi.org/10.3141/1807-12.
Molnar, C., 2019. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.
Nam, D., Kim, H., Cho, J., Jayakrishnan, R., 2017. A model based on deep learning for predicting travel mode choice. Transportation Research Board 96th Annual Meeting.
Omrani, H., 2015. Predicting travel mode of individuals by machine learning. Transp. Res. Proc. 10, 840–849. https://doi.org/10.1016/j.trpro.2015.09.037.
Paredes, M., Hemberg, E., O'Reilly, U.M., Zegras, C., 2017. Machine learning or discrete choice models for car ownership demand estimation and prediction? In: 5th IEEE Int. Conf. Model. Technol. Intell. Transp. Syst. MT-ITS 2017 - Proc. pp. 780–785. doi: 10.1109/MTITS.2017.8005618.
Parsa, A.B., Movahedi, A., Taghipour, H., Derrible, S., Mohammadian, A. (Kouros), 2020. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. Accid. Anal. Prev. 136, 105405. doi: 10.1016/j.aap.2019.105405.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 19, 29–33. https://doi.org/10.1145/2786984.2786995.
Pineda-Jaramillo, J., 2021. Travel time, trip frequency and motorised-vehicle ownership: a case study of travel behaviour of people with reduced mobility in Medellín. J. Transp. Heal. 22, 101110 https://doi.org/10.1016/J.JTH.2021.101110.
Rahman, H.Z., 2016. Bangladesh 2016 Politics, Governance and Middle Income Aspirations. Power and Participation Research Centre. Dhaka, Bangladesh.
Ribeiro, M.T., Singh, S., Guestrin, C., 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16. Association for Computing Machinery, New York, NY, USA, pp. 1135–1144. doi: 10.1145/2939672.2939778.
Salas, P., de la Fuente, R., Astroza, S., Carrasco, J.A., 2022. A systematic comparative evaluation of machine learning classifiers and discrete choice models for travel mode choice in the presence of response heterogeneity. Expert Syst. Appl. 193 https://doi.org/10.1016/j.eswa.2021.116253.
Sifringer, B., Lurkin, V., Alahi, A., 2020. Enhancing discrete choice models with representation learning. Transp. Res. Part B: Methodol. 140, 236–261. https://doi.org/10.1016/j.trb.2020.08.006.
Storchmann, K., 2005. Long-run gasoline demand for passenger cars: the role of income distribution. Energy Econ. 27, 25–58. https://doi.org/10.1016/j.eneco.2004.03.002.
Train, K.E., 2009. Discrete choice methods with simulation, second edition. Cambridge University Press. doi: 10.1017/CBO9780511753930.
TYPSA, 2019. Feasibility Study and Preliminary Design for construction of Dhaka Subway.
van Cranenburgh, S., Wang, S., Vij, A., Pereira, F., Walker, J., 2022. Choice modelling in the age of machine learning - discussion paper. J. Choice Model. https://doi.org/10.1016/j.jocm.2021.100340.
Wadud, Z., 2020. The effects of e-ridehailing on motorcycle ownership in an emerging-country megacity. Transp. Res. Part A: Policy Pract. 137, 301–312. https://doi.org/10.1016/j.tra.2020.05.002.
Walker, J., Vij, A., Brathwaite, T., 2019. Choice modelling in an age of machine learning. In: International Choice Modelling Conference (ICMC) Kobe.
Wang, S., Mo, B., Zhao, J., 2020a. Deep neural networks for choice analysis: architecture design with alternative-specific utility functions. Transp. Res. Part C: Emerg. Technol. 112, 234–251. https://doi.org/10.1016/j.trc.2020.01.012.
Wang, F., Ross, C.L., 2018. Machine learning travel mode choices: comparing the performance of an extreme gradient boosting model with a multinomial logit model. Transp. Res. Rec. 2672, 35–45. https://doi.org/10.1177/0361198118773556.
Wang, S., Wang, Q., Zhao, J., 2020b. Deep Neural networks for choice analysis: extracting complete economic information for interpretation. Transp. Res. Part C: Emerg. Technol. 118, 102701 https://doi.org/10.1016/j.trc.2020.102701.
Wang, S., Wang, Q., Zhao, J., 2020c. Multitask learning deep neural networks to combine revealed and stated preference data. J. Choice Model. 37 https://doi.org/10.1016/j.jocm.2020.100236.
Wen, X., Xie, Y., Wu, L., Jiang, L., 2021. Quantifying and comparing the effects of key risk factors on various types of roadway segment crashes with LightGBM and SHAP. Accid. Anal. Prev. 159, 106261 https://doi.org/10.1016/J.AAP.2021.106261.
World Bank, 2005. STP The Strategic Transport Plan For Dhaka.
Xie, C., Lu, J., Parkany, E., 2003. Work travel mode choice modeling with data mining: decision trees and neural networks. Transp. Res. Rec. 50–61. https://doi.org/10.3141/1854-06.
Yang, C., Chen, M., Yuan, Q., 2021. The application of XGBoost and SHAP to examining the factors in freight truck-related crashes: an exploratory analysis. Accid. Anal. Prev. 158, 106153 https://doi.org/10.1016/J.AAP.2021.106153.
Yao, R., Bekhor, S., 2020. Data-driven choice set generation and estimation of route choice models. Transp. Res. Part C: Emerg. Technol. 121 https://doi.org/10.1016/j.trc.2020.102832.
Zannat, K., Choudhury, C.F., Hess, S., 2021. Modeling departure time choice of car commuters in Dhaka, Bangladesh. Transp. Res. Rec. J. Transp. Res. Board, 036119812110398. https://doi.org/10.1177/03611981211039840.

Zegras, P., Gakenheimer, R., 2006. Driving Forces in Developing Cities' Transportation Systems: Insights from Selected Cases, Developing Country Urban Transport Cases, Massachusetts Institute of Technology.
Zhao, X., Yan, X., Van Hentenryck, P., 2019. Modeling Heterogeneity in Mode-Switching Behavior Under a Mobility-on-Demand Transit System: An Interpretable Machine Learning Approach. arxiv.org/abs/1902.02904.
Zhao, X., Yan, X., Yu, A., Van Hentenryck, P., 2020. Prediction and behavioral analysis of travel mode choice: a comparison of machine learning and logit models. Travel Behav. Soc. 20, 22–35. https://doi.org/10.1016/j.tbs.2020.02.003.