Titre: Title:	Who is more likely to buy electric vehicles?
Auteurs: Authors:	Hamed Naseri, Owen Waygood, Zachary Patterson, & Bobin Wang
Date:	2024
Type:	Article de revue / Article
Référence: Citation:	Naseri, H., Waygood, O., Patterson, Z., & Wang, B. (2024). Who is more likely to buy electric vehicles? Transport Policy, 155, 15-28. https://doi.org/10.1016/j.tranpol.2024.06.013

Document publié chez l'éditeur officiel
Document issued by the official publisher

Titre de la revue: Journal Title:	Transport Policy (vol. 155)
Maison d'édition: Publisher:	Elsevier
URL officiel: Official URL:	https://doi.org/10.1016/j.tranpol.2024.06.013
Mention légale: Legal notice:	

Transport Policy 155 (2024) 15–28

Contents lists available at ScienceDirect

Transport Policy

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

Who is more likely to buy electric vehicles?

Hamed Naseri [a,*], E.O.D. Waygood [a], Zachary Patterson [b], Bobin Wang [c]

[a] *Department of Civil, Geological, and Mining Engineering, Polytechnique Montréal, Montréal, QC, H3T 1J4, Canada*
[b] *Concordia Institute for Information Systems Engineering, Concordia University, Montréal, QC, H3G 1M8, Canada*
[c] *Department of Mechanical Engineering, Université Laval, Québec, QC, G1V 0A6, Canada*

A R T I C L E I N F O

Keywords:
Electric vehicle
Ensemble learning
Variables analysis
Accumulated local effects

A B S T R A C T

To promote electric vehicles, it is vital to know what impacts the preferences for electric vehicles over conventional fuel-based cars. To address this, a discrete choice experiment is developed and integrated into a survey. An online survey was conducted in Canada with 2062 valid responses. Different labels are designed for the survey to determine the most effective GHG information framing to increase the influence of such information on decisions. In this study, the influence of lifecycle emissions is considered. Three ensemble learning techniques are applied and they are compared based on prediction accuracy, and the most accurate technique is applied to determine the relative influence of variables on the intention to buy electric vehicles. Further, the interaction of variables is investigated using xgbfir. Subsequently, Accumulated Local Effect (ALE) is employed to examine the influence direction of top variables on the electric vehicle purchase likelihood. The results suggest that environmental attitudes and purchase price are the most influential parameters on the intention to buy electric vehicles. Moreover, those who are extremely worried about climate change, do not own a car, and self-identified as being at the top of the climate change stage of change are more likely to buy electric vehicles.

1. Introduction

The transportation industry is the second largest source of global greenhouse gas (GHG) emissions, and the GHG emissions of this industry tend to increase (Lamb et al., 2022). The transportation sector is responsible for 24% of global CO_2 emissions (Austmann, 2021), contributing to nearly 25% of international fossil fuel consumption (Jahanbakhsh et al., 2020). Hence, concerted efforts are required to reduce this sector's emissions. Many countries around the world are attempting to reduce transportation-based GHG emissions (Hardman, 2019). One possible solution to lower GHG emissions caused by the transportation sector and dependency on oil is to promote Electric Vehicles (EVs) (Kucukoglu et al., 2021).

Different policies have been made in many countries to promote EVs, such as subsidization of EVs (Dong, 2022; He et al., 2021; Lim et al., 2022), enhancing EVs' charging infrastructure (Funke et al., 2019), and taxing on fossil fuel (Barros et al., 2021; Colgan and Hinthorn, 2021). These implementations have been effective in a few countries. As an example, in Norway, 72.6% of new registered passenger vehicles in 2020 were battery electric vehicles or plug-in hybrid electric vehicles (Fevang et al., 2021). However, so far, such policies have not significantly increased the share of EVs on the vehicle market of many other countries, and only 1% of global new vehicle sales were EVs in 2020 (Austmann, 2021). In Canada, the share of EVs in the new vehicle market was 4% in 2020 (Long and Axsen, 2022). Thus, the parameters increasing the intention to buy EVs should be detected, and they should be better analyzed to promote EVs and increase their share in the vehicle market.

There are many parameters that can influence EV preferences, such as socio-demographic variables, individuals' attitudes, and GHG information presentation. Therefore, the influence of these variables on the intention to buy EVs should be tested. To this end, a discrete choice experiment should be designed, and a survey should be implemented to collect the data. Subsequently, an accurate technique should be applied for modeling in order to obtain accurate results. Machine learning methods are accurate prediction models, and they can be applied to predict who buys EVs. Nonetheless, they are black-box tools, and they cannot be easily applied to set new policies. Therefore, interpretation techniques can be applied to address this issue.

* Corresponding author. Department of Civil, Geological, and Mining Engineering, Ecole Polytechnique Montreal University, Montreal, Canada.
E-mail addresses: hamed.naseri@polymtl.ca (H. Naseri), owen.waygood@polymtl.ca (E.O.D. Waygood), zachary.patterson@concordia.ca (Z. Patterson), bobin.wang@gmc.ulaval.ca (B. Wang).

https://doi.org/10.1016/j.tranpol.2024.06.013
Received 13 March 2024; Received in revised form 30 May 2024; Accepted 17 June 2024
Available online 18 June 2024
0967-070X/

2. Literature review

Different studies have been conducted to determine who is more likely to buy EVs and which variables strongly impact the willingness to buy EVs. As such, Ling et al. (2021) investigated different socio-demographic variables and individuals' attitudes to evaluate who intended to buy EVs in China. The findings suggested that gender, inclination to ICEVs, household income, and previous EV driving experiences significantly impacted the likelihood of preferences for EVs. Further, males and those who had a higher household income had a stronger intention to buy EVs. Age, gender, residential location, and household type were found to be significant parameters for the intention to buy EVs in South Korea (Lashari et al., 2021). Further, males, aged over 30, capital area residents, and those living in apartments were more likely to buy EVs in South Korea.

Likewise, the influence of individuals' attitudes on intention to buy EVs has been tested. As such, Simsekoglu and Nayum (2019) examined the impacts of knowledge and psychological factors on the willingness to buy EVs. The results indicated that perceived behavioral control, subjective norms, and environmental attributes of EVs are directly linked to the intention to buy EVs. The attitudes of millennials toward buying EVs were also examined (Vafaei-Zadeh et al., 2022). For millennials, perceived behavioral control, subjective norms, attitude, and environmental self-image showed a significantly positive association with EV purchase intentions.

Some other studies investigated the impact of GHG information presentation on the willingness to for EVs. Daziano et al. (2017) applied normative and gain framings to better represent GHG information on car purchase labels. Accordingly, five labels were designed, and a discrete choice experiment was used to evaluate the effectiveness of the different labels. The results suggested that presenting GHG information with respect to a government objective could significantly increase the willingness to pay for GHG emissions. However, in that study, vehicles were not labelled as EVs or internal combustion engine vehicles (ICEVs). Continuing such research, Wang et al. (2021) designed six GHG information labels using gain, norm, and hedonic framings. In this experiment, vehicles were labelled as ICEV, plug-in hybrid EV (PHEV), and battery EV (BEV). In this study, the addition of hedonic information increased the influence of the government objective framing. However, those studies did not examine directly who would be more likely to choose EVs. Further, those studies were based on tail-pipe emissions, but considering that the climate crisis is a global environmental problem, the lifecycle emissions (i.e., all emissions associated with the production, use, and end-of-life) should be presented.

In terms of modeling, various techniques have been applied to investigate which factors influence the intention to buy EVs and how EV preferences can be increased, such as regression (Brase, 2019), maximum-likelihood estimation method (Li et al., 2013), Structural Equation Modelling (Krishnan and Koshy, 2021), discrete choice model (Plananska and Gamma, 2022), and machine learning (Sobiech-Grabka et al., 2022). Discrete choice models have been the widely-used techniques to investigate EV adoptions. Multinomial Logit and Mixed Logit are appropriate techniques to model discrete choice experiment data (Louviere et al., 2000). Although Multinomial Logit and Mixed Logit have been extensively used to model people's intention to buy EVs, several studies have demonstrated that Latent Class Models outperform Multinomial Logit and Mixed Logit when comparing model fit and accuracy (Boeri et al., 2020; Cerwick et al., 2014; Greene and Hensher, 2003; Li et al., 2019; Shen, 2009). Accordingly, some researchers have employed Latent Class Models to model the intention to buy EVs, e.g., Ferguson et al. (2018), Abotalebi et al. (2019), Plananska and Gamma (2022).

Although discrete choice models are easy to interpret and provide useful information on EV choice/preference, their prediction accuracy are often less than powerful machine learning techniques (Pham et al., 2022; Pineda-Jaramillo and Arbeláez-Arenas, 2022; Salas et al., 2022; Wang et al., 2020). Contrary to discrete choice models, machine learning methods have flexible structures, and they can capture non-linear relationship between the input (e.g., socio-demographic variables and vehicle attributes) and output variables (e.g., intention to buy EVs). Machine learning methods can apply unparalleled computation, and as a result, they can be used even for unstructured and high-dimensional data (Pi, 2021).

Therefore, machine learning methods have been receiving more attention in investigating EV preferences in recent years. Sobiech--Grabka et al. (2022) applied four machine learning algorithms, including Support Vector Machine, K-nearest neighbor, classification and regression trees, and Random Forest, to predict customers' intention to buy EVs in Poland. Random Forest, as an ensemble learning method, outperformed other machine learning methods regarding prediction accuracy and kappa coefficients. The results of Random Forest suggested that belief about the effectiveness of emissions of electricity production in Poland and agreement with the statement that ICEV will be replaced with EVs in ten years had the most significant relative influence on willingness to buy EVs. Hence, the individuals' vehicle engine preferences could be accurately predicted, and the variables with the strongest relative influence could be detected. However, the influence direction of top variables on intention to buy EVs could not be determined since machine learning techniques are black-box tools.

Moreover, using machine learning techniques, prediction methods have been recently developed to predict who selected EVs over ICEVs. As such, different machine learning techniques, such as Generalized Linear Model, Deep Learning, and Gradient Boosting, were applied to predict who buys electric vehicles (Bas et al., 2021b). In this regard, a survey was implemented in the State of Maryland, United States, and 374 respondents joined the survey. Since their study only considered one state, the results might not be highly generalizable. The study's findings revealed that Gradient Boosting was the best prediction algorithm to predict EV buyers when comparing the prediction accuracy. Subsequently, Gradient Boosting was used in a further analysis showing that the price of EVs, range of EVs, preference of next vehicle engine type, county, environmental concern, and attitude toward EVs could impact willingness to adopt electric vehicles. Finally, Local Interpretable Model-Agnostic Explanation (LIME) was used to make the results of Gradient Boosting (as a black-box technique) interpretable. Nonetheless, LIME is a method for interpretability at the individual level, and it could not detect the overall influence direction of significant variables.

Another prediction model was developed to predict EV buyers in the State of Maryland, United States using the same dataset (Bas et al., 2021a). Different machine learning techniques, i.e., Gradient Boosting, Random Forest, Artificial Neural Networks, Support Vector Machines, and Deep Neural Networks, were applied for the prediction process. The results of the most accurate model (Support Vector Machines) found that environmental concerns, household income, and attitudes toward EVs had the strongest influence on EV adoption. The top variables could be detected; however, how these variables impact the intention to buy EVs could not be investigated due to the black-box nature of machine learning techniques.

As can be seen from the above studies, prediction models have been generated to determine who prefers EVs over ICEVs. In those models, different types of EVs (i.e., Battery Electric Vehicles (BEV) and Plug-in Hybrid Electric Vehicles (PHEV)) were considered in the same category. However, these two types of vehicles have various characteristics such as range, emission profiles, and cost. Previous research has demonstrated that BEV-oriented and PHEV-oriented are different groups in Canada in terms of attitudes, socio-demographic variables, and level of concern about the environment (Ferguson et al., 2018). Therefore, separate prediction models for different types of EVs are required to accurately predict EV buyers with different orientations. Further, the influence of GHG information framing on individuals' willingness to buy EVs has not been tested in EV buyer prediction models.

Regarding modeling, machine learning methods can accurately predict who buys EVs. However, they are black-box tools, and they cannot be easily applied to set new policies. Ensemble learning techniques (e.g., Gradient Boosting Decision Tree (Dong et al., 2022), Random Forest (Naseri et al., 2021a), and AdaBoost (Naseri et al., 2021b)) are generally applied to solve this problem because these techniques can rank the variables based on their relative influence on the problem's response variable. That is, ensemble learning techniques can detect the variables that strongly influence the intention to buy EVs (Bas et al., 2021b). However, they cannot illustrate the influence direction (e.g., positively, negatively, linearly, or non-linearly) of each variable on the response variable.

To solve the mentioned problems, this study aims to develop new prediction models to investigate who prefers EVs over ICEVs, and which variables impact the willingness to buy EVs. Eight vehicle labels are tested, including six new vehicle labels, to examine the influence of GHG information framing on individual choices and detect the most effective labels to maximize the intention to buy EVs. Seven of those labels present lifecycle emissions. A Discrete Choice Experiment (DCE) is designed to conduct a survey. Two separate prediction models are developed for BEVs and PHEVs to predict Canadians' vehicle engine choices and better understand BEV-oriented and PHEV-oriented groups. Different ensemble learning methods are applied for the modeling process since they can rank variables based on their relative influence on the intention to buy BEVs and PHEVs. Further, the interaction of variables is investigated, and the top interactions are investigated to better understand who is more likely to buy BEVs and PHEVs. Since ensemble learning methods can not present how top variables impact the intention to buy EVs, Accumulated Local Effects (ALE) is applied to interpret the results and illustrate the influence direction of top variables on EV preferences.

The contributions of this study are as follows.

- Designing new vehicle labels (by integrating the current consequences of climate change that Canadians experience) to investigate the impact of GHG information presentation on EV preferences and determine the most effective labels for EV promotion.
- Applying life-cycle CO_2 emissions in vehicle labels rather than tailpipe emissions.
- Detecting the most accurate technique to predict Canadians' vehicle engine choices by comparing robust prediction techniques.
- Prioritizing the variables based on their relative influence on the intention to buy BEVs and PHEVs.
- Examining the interaction of variables on EV choice.
- Using a powerful interpretation model (ALE) to evaluate the probability of choosing BEVs and PHEVs over ICEVs in different circumstances and for different populations.

3. Methodology

The purpose of this study is to detect who is more likely to buy EVs and which variables strongly affect the willingness to buy EVs. In this regard, a DCE is designed, according to Wang et al. (2021), to conduct a survey. Then, the data is collected using an online survey. Subsequently, data is prepared for the modeling process, and ensemble learning methods are used for modeling. Then, the variables will be ranked based on their relative influence on the intention to buy EVs. Subsequently, variables interactions will be investigated, and top interactions will be detected. Finally, ALE will be applied to represent how the strongest variables and the strongest variables interactions influence the probability of buying EVs.

In this part, the experiment design and data collection process are explained. Afterward, the methods applied for modeling are presented. Ultimately, the technique used for results interpretation is presented.

3.1. Survey and data collection

A Discrete Choice Experiment (DCE) was designed to conduct a survey and collect the required data. An online survey was used to collect responses in the Fall of 2022. The respondents were recruited by Survey Engine to participate in the online survey. A pan-Canadian sample was chosen because attitudes toward climate change and support for government actions vary in different provinces of Canada and previous research has shown geographic variances in EV intention. The participants were British Columbia (BC), Alberta (AB), Saskatchewan (SK), Manitoba (MB), Ontario (ON), and Quebec (QC). As the survey was on vehicle purchase, respondents had to currently have a driver's license and be 18 years or older. The survey was offered in two languages (English and French). Some trap questions were embedded in different survey phases to identify respondents who were not paying attention to all questions. Such participants were excluded from the survey, and the remaining ones were considered in the final dataset. The final dataset included 2062 participants from the 2400 who completed the survey. In this section, the attitudes of participants and their socio-demographics are first introduced. The designed DCE is then presented.

3.1.1. Socio-demographic and environmental attitudes

Table 1 indicates the respondents' information in terms of socio-demographic variables and environmental concern. The average age of respondents was 49.7 years, 17% were francophones, 48.9% were females, and 91% owned a vehicle when filling out a survey. Climate Change Stage of Change (CC-SoC) and level of worried about climate change were applied to assess individuals' attitudes toward environment and climate change. Roughly 15% of respondents were extremely worried about the climate change. CC-SoC is a powerful index to capture attitudes and behavior with respect to personal climate emissions (Waygood et al., 2021). The respondents could select between one of the five following options to evaluate their CC-SoC.

CC-SoC1: I am not concerned about climate change.
CC-SoC2: I am concerned about climate change, but I do not plan to reduce my emissions.
CC-SoC3: I would like to reduce my emissions, but I don't know how.
CC-SoC4: I would like to reduce my emissions and will do so in the future.
CC-SoC5: I have already reduced my emissions significantly.

It was demonstrated that CC-SoC can strongly evaluate people's environmental motivations, and it can suitably be replaced with more complex environmental measures (Waygood et al., 2021). Moreover, the participants from Quebec (the primarily francophone province) were divided into two groups based on their language to capture the impact of language on individuals' choices. Those two groups were English Quebecers (QC-English) and French Quebecers (QC-French).

3.1.2. Discrete choice experiment (DCE)

This study applied a DCE to investigate which variables significantly impact EV preference. The DCE was presented before socio-demographic questions in the survey. For half of the respondents, the climate change attitude question (i.e., CC-SoC) was asked before the DCE, and the others responded to the CC-SoC question after DCE to test any possible impact of this question on individuals' choices. Different choice situations (hypothetical tasks) were presented to respondents by DCE. The experimental design presented the choices' attributes, and respondents should choose between the presented alternatives according to their preferences.

In this research, the DCE was designed based on the details presented by Wang et al. (2021), as it had the same underlying experiment and precise econometric model. In the designed DCE, each respondent needed to do 12 different choice tasks. In six tasks, respondents should select between a BEV and an ICEV. In the remaining six tasks, they needed to choose between a PHEV and an ICEV. As mentioned, 2062 participants completed the survey without missing the trap questions (that test whether the respondent is paying attention). Therefore, the

Table 1
Select socio-demographic and environmental attitudes of respondents.

Variable	Frequency	Variable	Frequency
Gender		CC-SoC	
Male	1045	CC-SoC1	176
Female	1008	CC-SoC2	234
Other	9	CC-SoC3	333
Education attainment		CC-SoC4	809
No formal education	2	CC-SoC5	510
Elementary school education	1	Province and language	
Less than high school equivalent	24	BC	300
High school diploma or equivalent	325	AB	265
Registered Apprenticeship or other trades certificate or diploma	108	SK	64
College, CEGEP or other non-university certificate or diploma	454	MB	94
University bachelor's degree	770	ON	903
Degree in medicine, dentistry, veterinary medicine or optometry (M.D., D.D.S., D.M.D., D.V.M., O.D.)	26	QC-English	86
Master's degree (e.g., M.A., M.Sc., M.Ed., M.B.A.)	294	QC-French	350
Doctoral degree (e.g., Ph.D.)	58	Employment	
Household income		Full-time	1068
Less than $5000	12	Part-time	191
$5000-$7499	8	Homemaker	39
$7500-$9999	7	Full-time student	72
$10,000-$12,499	9	Retired	600
$12,500-$14,999	10	Not currently employed, but looking for work	46
$15,000-$19,999	18	Other	47
$20,000-$24,999	36	Ethnicity[a]	
$25,000-$29,999	34	North American Aboriginal	36
$30,000-$34,999	40	Other North American	388
$35,000-$39,999	53	European	1083
$40,000-$49,999	105	Latin, Central and	27
$50,000-$59,999	138	South American	18
$60,000-$74,999	190	African	16
$75,000-$84,999	128	Asian	354
$85,000-$99,999	173	Oceania	6
$100,000-$124,999	241	Prefer not to answer	134
$125,000-$149,999	157	Age	
$150,000-$174,999	120	Mean	49.69
$175,000-$199,999	83	How worried are you about climate change?	
$200,000 or more	166	Not at all worried	182
Prefer not to answer	334	Slightly worried	443
Currently own a vehicle		Somewhat worried	589
Yes	1877	Very worried	533
No	185	Extremely worried	315

[a] Canadian Census approach was applied to ask ethnicity questions. For more details, please visit https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E.

number of completed tasks for each EV type (BEV and PHEV) was 2062 $\times$ 6 = 12,372. The BEV-related tasks were applied to examine which variables influence BEV purchase likelihood, and PHEV-related tasks were utilized to investigate which variables impact PHEV purchase likelihood.

The choice attributes that were taken into account in the DCE design are: vehicles' price, range of electric vehicles, annual fuel or electricity cost, and life-cycle CO_2 emissions. In each choice, these choice attributes were changed, and the respondents needed to select between choices in a new choice situation. That is, the survey participants should select between an EV label and an ICEV label 12 times, and in each choice scenario, the attributes of EVs and ICEVs varied. For calculating the annual fuel or electricity cost, a yearly driven distance of 20,000 km was considered based on the current Natural Resources Canada (NRCan) label. The NRCan label is the current label applied in Canada. A D-efficient design was adopted in Ngene for DCE in this study. The attributes' levels applied for designing DCE are shown in Table 2.

Attributes of the vehicles are presented in the survey using vehicle labels. Since individual vehicle engine choices are significantly influenced by GHG information framing (Daziano et al., 2017), eight labels are used to present CO_2 emissions. To gain additional information on GHG information framing, please read Wang et al. (2021). The applied labels in the survey are shown in Fig. 1. These labels are NRCan label (the current label applies in Canada), NRCan label with life cycle emissions, emojis, different layout of label, flood, fire, disaster, and leaves. The NRCan label is the current mock-up in Canada, and it represents tailpipe emissions of different vehicles since the GHG information. However, several studies demonstrated that tailpipe emission does not correctly represent the overall GHG emissions of vehicles as a significant amount of GHG is generated in production, battery production, operation, maintenance, and disposal (Ambrose et al., 2020; Kosai et al., 2022).

Therefore, life cycle emissions of vehicles are used to design second to eight treatments (i.e., labels). The life cycle emissions of vehicles are calculated based on the details provided by De Souza et al. (2018). The second label (i.e., NRCan label with life cycle emissions) is the same as the first label but uses the life cycle emissions rather than tailpipe emissions. Emojis were found to be the most effective label for promoting EVs in the previous study (Xun et al., 2022).

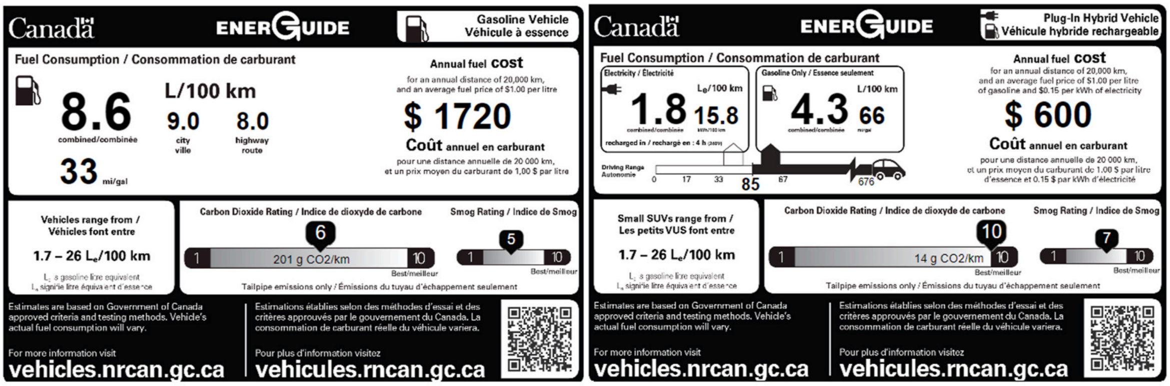
Several studies showed that climate change information should be presented using its negative impacts on your neighbourhood and the way it influences people like you (McDonald et al., 2015). In Canada, devastating floods and widespread wildfires are the current consequences of climate change that Canadians experience (Government of Canada, 2021). Therefore, the following framings (i.e., flood, fire, disaster, and leaves) are designed based on the current disasters in Canada. By 2030, the Canadian government aims to reduce its 2005 GHG emissions by 30% (NRCan, 2019). Accordingly, in Treatment 3 and 8, the emissions are framed with respect to the 30% GHG reduction goal. For more information on framing design with the 30% GHG reduction, please read Wang et al. (2021).

3.2. Problem modeling

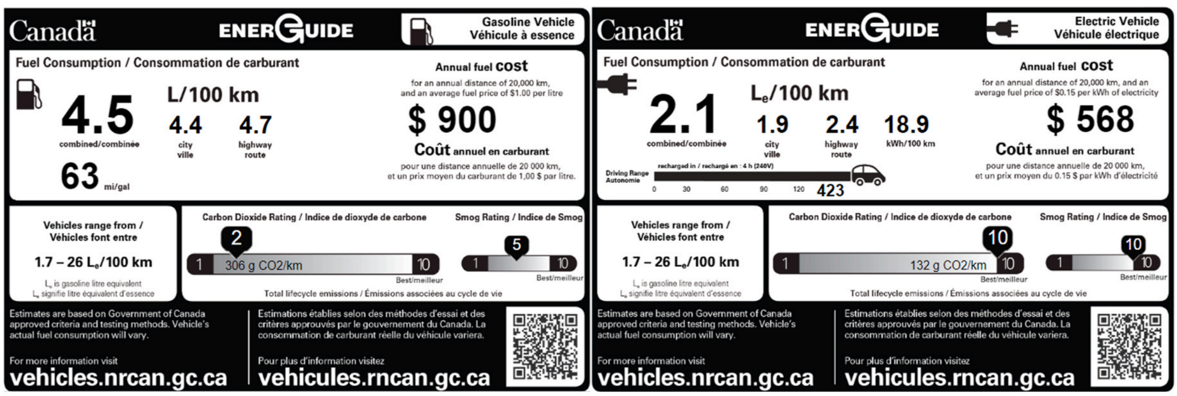
After data collection, both datasets (BEV and PHEV) were randomly divided into two groups: training data (80% of total samples) and testing data (20% of total samples). Three ensemble learning methods were employed to predict individuals' decisions regarding choice tasks and determine which variables influence EV preference. In other words, eXtreme Gradient Boosting (XGB), Light Gradient Boosting Machine

Table 2
The levels of attributes used in DCE.

Alternatives	Price (CA$)	Monthly electricity or fuel cost (CA$)	CO_2 emissions (kg/month)	EV battery range (km)
ICEV	[22000; 26000; 30000]	[900; 1420; 1720]	[210; 270; 306]	–
PHEV	[30000; 38000; 48000]	[600; 840]	[168; 186]	[40; 85; 203]
BEV	[40000; 48000; 56000]	[444; 568]	[132; 150]	[240; 423; 600]



(a) Treatment 1: NRCan label with tailpipe emission (current label in Canada)



(b) Treatment 2: NRCan label with life cycle emissions

Fig. 1. The current and new designed labels to determine the efficient treatment for better present GHG emissions.

(LGBM), and CatBoost were used to predict individual choices and investigate the relative influence of variables on intention to buy EVs. For each technique, K-fold cross-validation (considering K = 5) and the Randomized Search were used in python's SciPy library to tune hyper-parameters. Since Randomized Search is a heuristic-based tuning method, it was run 30 times (according to (Naseri et al., 2022)) for each ensemble technique. Then, the best hyperparameter set was employed to generate the final prediction models.

The models' input variables were CC-SoC, worried about climate change, age, province and language, education, gender, car ownership, employment status, ethnicity, household income, political spectrum, the price ratio of EV to ICEV, emission ratio of EV to ICEV, EV range, fuel cost ratio of EV to ICEV, and the type of label (treatment) applied for presenting GHG information. After running each technique (i.e., XGB, LGBM, and CatBoost), the most accurate technique for each dataset was determined by comparing testing data accuracy and testing data F1-score. Subsequently, the most accurate technique was employed to identify the relative influence of input variables on the response variables (i.e., task decisions). Then, python's xgbfir library was employed to determine the interaction of variables with the highest impact on individuals' choices. Xgbfir is a version of XGB that can detect the most important variables interactions according to the split point in the structure of XGB. This method was implemented based on the details provided by Hillel et al. (2019). Finally, ALE was utilized to understand the influence direction of top variables and top variables interactions on the probability of buying an EV. In the following sections, XGB, LGBM, CatBoost, and ALE are briefly explained.

3.2.1. eXtreme gradient boosting (XGB)

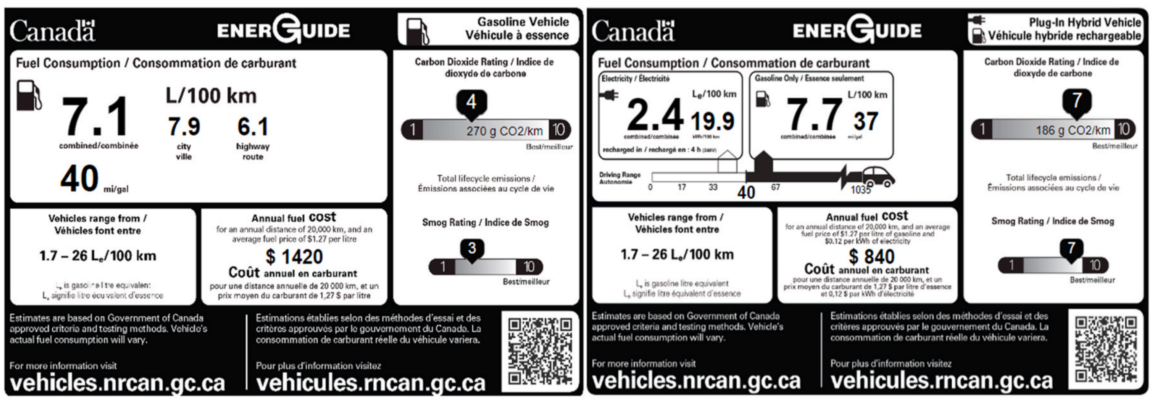
XGB is a powerful ensemble technique, which has been widely applied for modeling classification and regression problems. XGB applies rapid learning and parallel processing in the modeling process (Jeon et al., 2020). This technique can generally obtain high accuracy in problems with high complexity (Chen and Guestrin, 2016). This technique outperformed different machine learning techniques when comparing the prediction accuracy, such as Artificial Neural Network, Random Forest (Kim, 2021), Support Vector Machine (Nguyen-Sy et al., 2020), Logistic Regression (Wang and Sherry Ni, 2019), and Decision Tree (Jamal et al., 2021).

Since XGB is an ensemble learning technique, it uses different weak learners (i.e., Decision Trees) and combines them to generate a powerful prediction method. This is an iterative process, and an optimization process performs to maximize the prediction accuracy in each iteration. The optimization process aims to optimize the structures of weak learners using the differences between training samples and their corresponding predicted values in each iteration. To solve the optimization problems (i.e., optimize the XGB's structure), first- and second-order gradients are employed. Further, a regularized term is added to the optimization problem's objective function to reduce the probability of over-fitting (Zhu and Zhu, 2019).

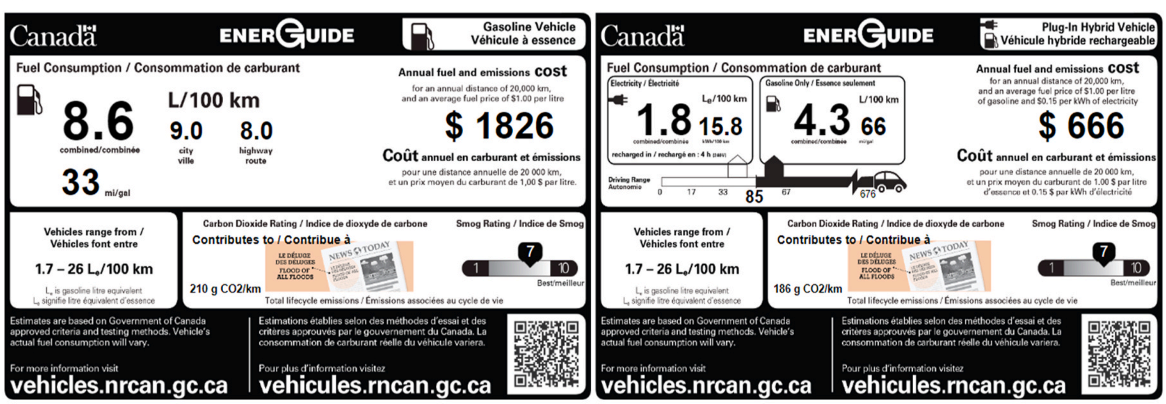
3.2.2. Light Gradient Boosting Machine (LGBM)

Light Gradient Boosting Machine (LGBM) is a quick and robust ensemble learning algorithm. This algorithm is generally used for prediction and determining the relative influence of variables on the output variable. LGBM is a new version of ensemble gradient boosting techniques, and it was introduced by Microsoft. Similar to XGB, LGBM uses a given number of weak prediction techniques to generate an accurate prediction model (Liu et al., 2020). LGBM uses parallel learning, and hence it can significantly reduce memory usage. It is a fast technique and

(c) Treatment 3: emojis



(d) Treatment 4: different layout of label



(e) Treatment 5: flood

Fig. 1. (*continued*).

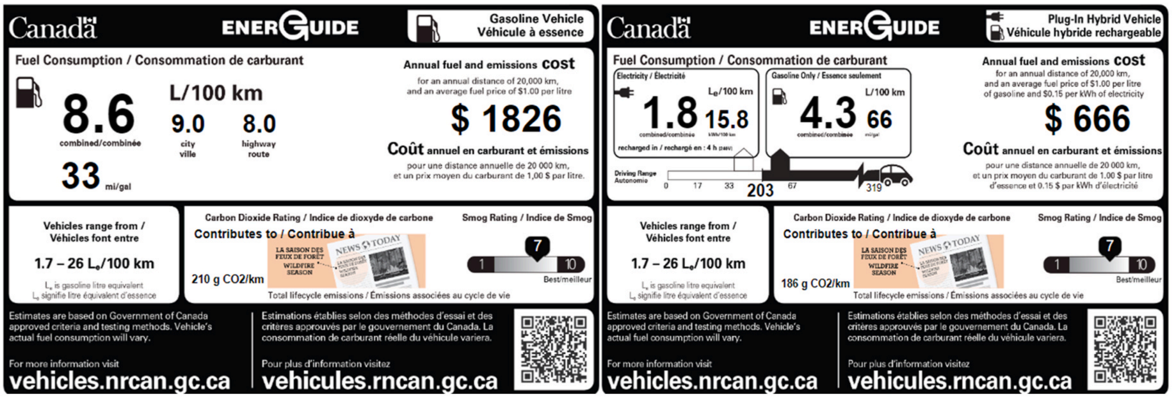
efficient for complicated prediction problems.

LGBM utilizes a leaf-wise leaf growth strategy limiting excessive depth growth in forming trees. The same leaves' layer can be split using the leaf-wise leaf growth strategy, and a multi-threaded optimization can be performed in LGBM. As a result, the model's complexity is automatically reduced, and overfitting is less likely to happen (Zhou et al., 2021). The results of previous studies showed that LGBM can outperform various prediction techniques in terms of prediction accuracy and running time, e.g., Back-Propagation Neural Network, Recurrent Neural Network, Stochastic Gradient Boosting (Cai et al., 2022), Random Forest (Tutica et al., 2021), CatBoost (Guo et al., 2020), K-Nearest Neighbors (Liu et al., 2020), Multi-layered Perceptron, Support Vector Machine, and XGB (Shangguan et al., 2022).

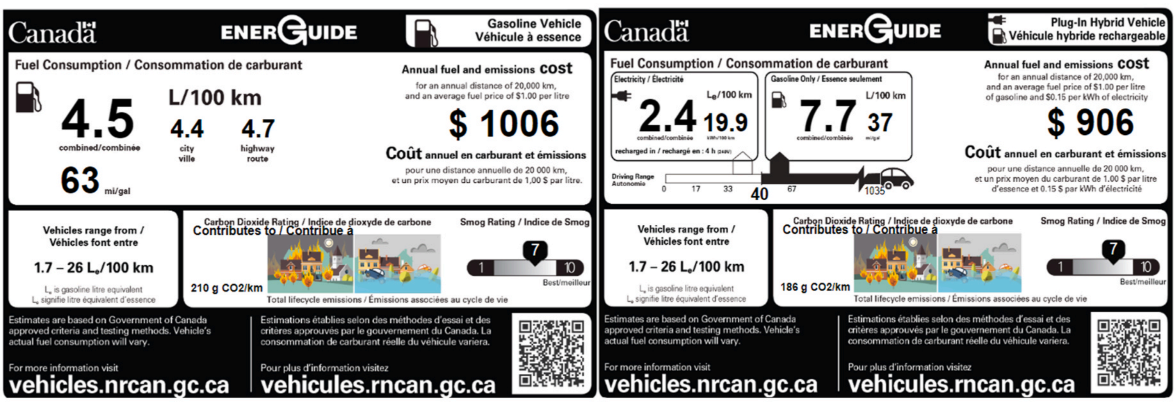
3.2.3. *CatBoost*

CatBoost is another machine learning method based on gradient boosting. This method can be used for modeling classification, regression, and ordinal variable prediction problems. Bayesian estimators are generally used in CatBoost to avoid overfitting (Dhananjay and Sivaraman, 2021). In CatBoost, categorical variables are replaced with binary variables to control computational costs in all iterations. By using efficient target-based statistics in CatBoost, categorical input variables can be directly modeled, thus reducing the running time (Hussain et al., 2021). Contrary to XGB and LGBM which use asymmetric decision tree structure, CatBoost applies symmetric (balanced) tree structures in the splitting process, being consistent across all nodes at the same depth of the tree (Prabhavathi et al., 2022).

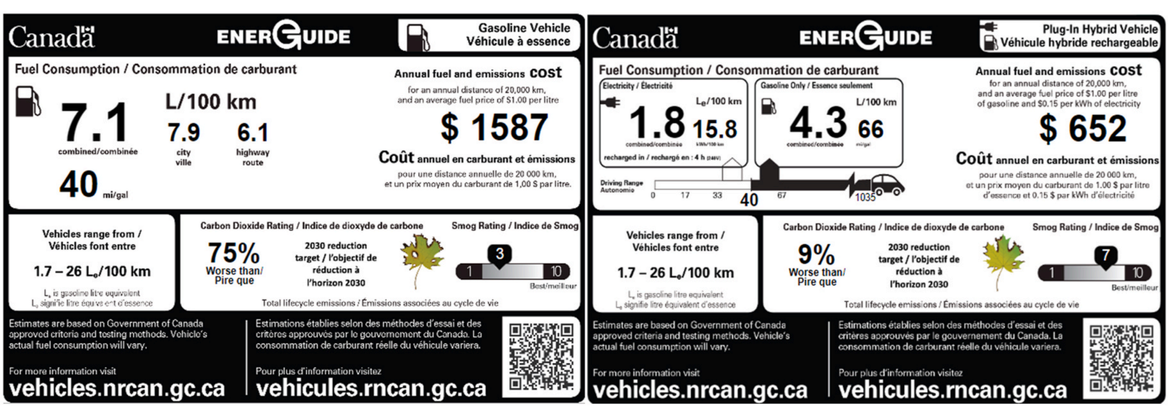
In this study, CatBoost is considered one of the applied techniques



(f) Treatment6: fire



(g) Treatment7: disaster



(h) Treatment 8: leaves

Fig. 1. (*continued*).

since it can present the relative influence of variables on the intention to buy EVs. Moreover, it was demonstrated that CatBoost is more accurate than conventional machine learning techniques, such as Gaussian Naïve Bayes, Decision Tree Classifier, Multi-layered Perceptron, Gradient Boosting Classifier, AdaBoost (Ibrahim et al., 2020), Long Short-Term Memory, Seq2seq, and Random Forest (Niu et al., 2021).

3.3. Results interpretation

The results of machine learning techniques should be interpreted for policy purposes in the transportation sector. Ensemble learning methods can sort variables according to the impact on the EV purchase likelihood. However, they can not illustrate how top ranked variables influence the EV purchase likelihood. To solve this issue, ALE is employed in this study to interpret the result of the most accurate classifier. That is, machine learning techniques are run, and they are compared based on their prediction accuracy. Then, the most accurate technique is determined. Subsequently, variables are ranked based on the relative influence presented by the most accurate classifier. Finally, ALE is used to illustrate the relation between the top ranked variables and the response variable (i.e., the EV purchase likelihood).

The ALE method was developed by Apley and Zhu (2020) to interpret black-box prediction tools. A set interval is used to divide independent variables in ALE. Afterward, the interval's upper and lower bounds are evaluated. ALE accumulates the estimated differences, and the average prediction is centered on zero (Kim et al., 2021). Hence, the probabilities of changing from a class to another could be estimated using the ALE plot. For more information about ALE, please read Apley and Zhu (2020).

4. Results and discussions

As mentioned, this study attempts to predict who prefers EVs to ICEVs, and determine which parameters have the highest contribution to increasing the intention to buy EVs. To this end, three prediction methods are applied to the prediction process. Then, ALE is used to interpret the results of the most accurate prediction model. In this section, the results of machine learning methods are first presented. Then, the outcomes of the most accurate classifier are discussed. Finally, ALE results are indicated to understand how variables impact the intention to buy EVs.

4.1. Performance of machine learning techniques

As discussed, the hyperparameters of ensemble learning techniques are tuned simultaneously using K-fold cross-validation and Randomized Search. The optimal hyperparameter values for both prediction problems and three prediction techniques are shown in Table 3. In this table, the accuracy score implies the average validation data accuracy over five folds. The learning rate is the shrinkage rate at each step of gradient descent and the maximum depth represents the maximum depth of decision trees in the ensemble structure. The minimum data in leaves denotes the minimum number of data observations at a leaf node, and the number of estimators the number of decision trees applies to form the ensemble structure.

Consequently, all prediction techniques for both prediction problems (i.e., the intention to buy BEVs and the intention to buy PHEVs) are run using their optimal hyperparameter values. Then, the testing data is used to evaluate the prediction power of the model and detect the most accurate classifier for each prediction problem. The performance of machine learning techniques for both prediction problems is shown in Table 4. As can be perceived, LGBM outperforms other techniques in terms of testing data accuracy for predicting the intention to buy BEVs. That is, the prediction accuracy of LGBM is 0.4% and 0.4% higher than XGB and CatBoost for predicting BEVs preferences over ICEVs. However, XGB prevails in LGBM and CatBoost in terms of predicting the intention to buy PHEVs. The accuracy of XGB to predict the intention to buy PHEVs is 0.6% and 0.6% higher than LGBM and CatBoost, respectively.

Running time is an essential index to compare the computational cost of soft computing techniques (Naseri et al., 2021a). As shown in Table 4, LGBM is much faster than XGB and CatBoost in both prediction problems. LGBM can reduce the running time of predicting the intention to buy BEVs by 77.6% and 92.8% than XGB and CatBoost, in the order given. Likewise, the running time of LGBM is 86.4% and 90.5% less than that of XGB and CatBoost when predicting the intention to buy PHEVs. Since the aim of this study is to present the results with the highest accuracy, LGBM is used for analyzing the individuals' intention to buy BEVs. Moreover, XGB is applied to investigate the variables influencing the intention to buy PHEVs.

Table 3
The optimal hyperparameters value for both prediction problems.

Hyperparameters	BEV vs ICEV			PHEV vs ICEV		
	XGB	LGBM	CatBoost	XGB	LGBM	CatBoost
Accuracy score	0.805	0.806	0.805	0.798	0.791	0.791
Learning rate	0.078	0.441	0.158	0.216	0.274	0.144
Maximum depth	12	12	8	6	22	8
Minimum data in leaves	11	8	36	51	40	39
Number of estimators	290	466	444	324	410	321

Table 4
The optimal hyperparameters value for both prediction problems.

Hyperparameters	BEV vs ICEV			PHEV vs ICEV		
	XGB	LGBM	CatBoost	XGB	LGBM	CatBoost
Testing data accuracy	0.823	0.827	0.823	79.7	79.1	79.1
Average running time (s)	48.3	10.84	150.84	63.85	8.66	91.49

4.2. The contribution of variables to the intention to buy EVs

In this section, the relative influence of variables on the intention to buy BEVs and PHEVs is presented. For each EV engine type (BEV and PHEV), the most accurate classifier is performed, and the importance weight of variables on the response variable is determined. Similarly, this process is performed for variables interactions using xgbfir. In this part, the relative influence of variables and their interactions on the intention to buy BEVs are first presented. Then, the importance wright of variables and their interactions on the willingness to buy PHEVs are represented.

4.2.1. The top variables on the intention to buy BEVs

As mentioned, LGBM (the most accurate method) is used to determine the relative influence of variables on the intention to buy BEVs. Those relative influences are indicated in Fig. 2. As can be seen, worried about climate change has the greatest impact on the intention to buy BEVs, with a relative influence of 11.9%. car ownership, price ratio of BEV to ICEV, and CC-SoC are the next leading variables with a relative influence of between 8% and 9%. Hence, it can be postulated that attitudes toward climate change (i.e., worried about climate change and CC-SoC) strongly influence BEV adoption. Regarding vehicle attitudes, the purchase price ratio of BEV to ICEV is the most important variable. On the other hand, emission ratio of BEV to ICEV, electricity range, and fuel cost ratio of BEV to ICEV are the three variables with minimum relative influence on the intention to buy BEVs. GHG information framing (i.e., treatment) is the 12th most impactful variable on vehicle choice, with a relative influence of 5.4%. Interestingly, the relative influence of GHG information framing on BEV preferences is higher than some vehicle attitudes, including emission ratio of BEV to ICEV, electricity range, and

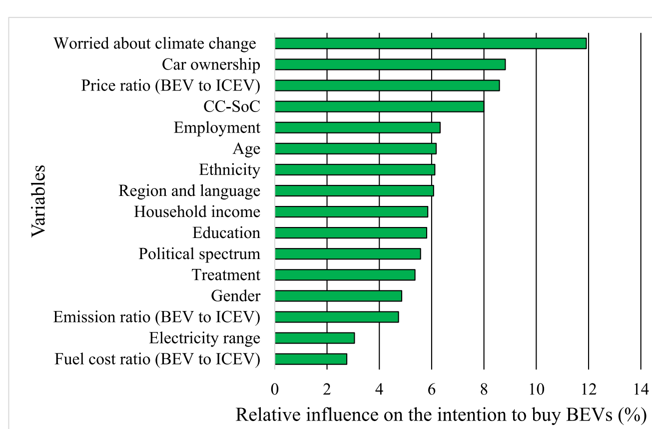


Fig. 2. The relative influence of variables on the intention to buy BEVs.

fuel cost ratio of BEV to ICEV. Therefore, it can be deduced that GHG information framing can be even more influential than the emission reduction of BEVs in terms of affecting the intention to buy BEVs.

The relative influence of variables interactions on the willingness to buy BEVs is examined using xgbfir. The relative influence of the top ten variables interactions is presented in Fig. 3. The interaction between worried about climate change and the purchase price ratio of BEV to ICEV is the leading interaction with respect to the willingness of BEV purchase. The contribution of this interaction on the intention to buy BEVs is 10.5%, which is 5.5% higher than the second top variable interaction. Therefore, it is recommended to investigate the worried about climate change and purchase price ratio when analyzing the intention to buy BEVs.

4.2.2. The top variables on the intention to buy PHEVs

XGB (as the most accurate method) is applied to evaluate the relative influence of variables on the intention to buy PHEVs, and the relative influences are represented in Fig. 4. Purchase price ratio of PHEV to ICEV is the most influential variable on the willingness to buy PHEVs, with an importance weight of 16.6%, 7.5% higher than the second influential variable. The second and third leading variables are worried about climate change and CC-SoC, with a relative influence of 9.1% and 8%, respectively. In the BEV model, the leading variable is worried about climate change while in the PHEV model, price ratio has the greatest relative influence. Accordingly, price plays a more crucial role in the PHEV preference than environmental attitude. The following variables are CC-SoC, car ownership, and education. Similar to the BEV model, emissions ratio of EV to ICEV, electricity range, and fuel cost ratio of EV to ICEV are the variables with the minimum relative influence on the willingness to buy PHEVs. GHG information framing is the eleventh top variable, with an importance weight of 5.4%. GHG information framing influences some other socio-demographic variables and vehicle attributes. Hence, it is vital to investigate the framings applied for present GHG information, and optimal treatments should be used to increase the intention to buy electric vehicles.

The relative influence of variables interactions on the intention to buy PHEVs is evaluated and shown in Fig. 5. The interaction between CC-SoC and the purchase price ratio is the leading interaction regarding the relative influence on the intention to buy PHEVs. The second top interaction is related to the interaction between worried about climate change and the purchase price ratio of PHEV to ICEV. Therefore, it can be postulated that purchase price and environmental attitude of individuals should be simultaneously analyzed to determine who is more likely to buy PHEVs.

4.3. Results interpretation

The top influential variables on the intention to buy BEVs and PHEVs

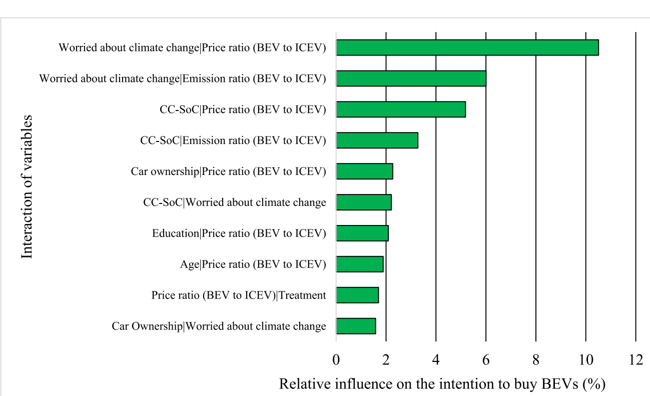


Fig. 3. The relative influence of top variables interactions on the intention to buy BEVs.

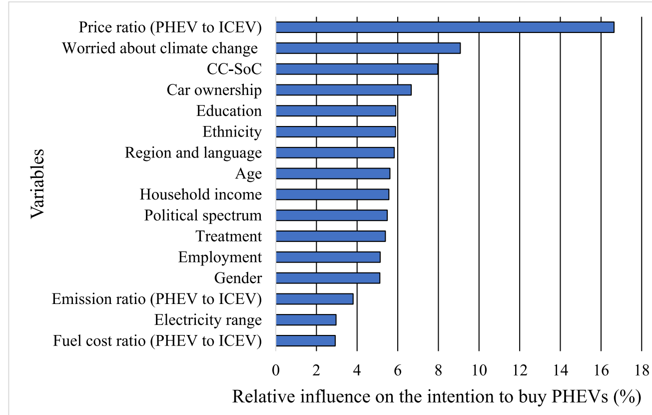


Fig. 4. The relative influence of variables on the intention to buy BEVs.

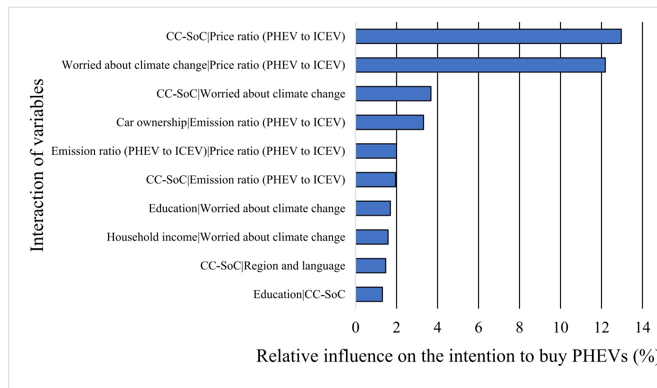


Fig. 5. The relative influence of top variables interactions on the intention to buy PHEVs.

can be determined using ensemble learning methods. However, these methods can not represent the influence direction of those variables on the response variable. To overcome this issue, ALE is employed to illustrate how top variables and top variables interactions impact the willingness to buy EVs. In ALE plots, the right axis and bar charts indicate the number of data samples in each group. The left axis and trend curve denote the difference in the probability of EV purchase between different groups. In this part, initially, the ALE results for top variables are presented. Afterward, the impact of top variable interaction on the intention to buy EVs is discussed.

4.3.1. Influence of top variables on EV purchase likelihood

Since purchase price ratio of EV to ICEV, worried about climate change, car ownership, and CC-SoC are within the top-five variables on the intention to buy both BEVs and PHEVs, these variables are investigated using ALE. Further, the influence of GHG information framing on the willingness to buy EVs is examined to determine the optimal vehicle label in Canada as it is within the control of policy makers.

4.3.1.1. Price. The influence of BEV and PHEV to ICEV price ratio on the willingness to buy EVs is illustrated in Figs. 6 and 7. In ALE plots, the bar charts signify the number of data observations in each category. As such, in Fig. 6, in approximately 1000 choices, the BEV to ICEV purchase price ratio was 1.33. In ALE plots, the line illustrates the probability of EV preference, and the likelihood values can be seen on the left axis. The probability of EV preference for two categories are compared by subtraction of their corresponding EV preference probabilities. As an example, the EV preference probability is 12% (14-2 = 12) reduced when the BEV to ICEV price ratio is increased from 1.33 to 1.54.

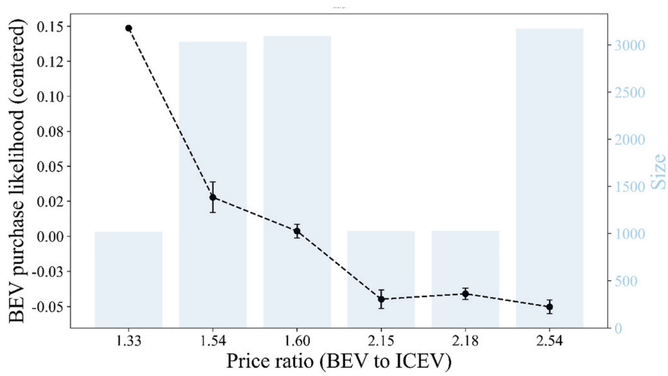


Fig. 6. The influence of BEV to ICEV price ratio on the likelihood of buying a BEV.

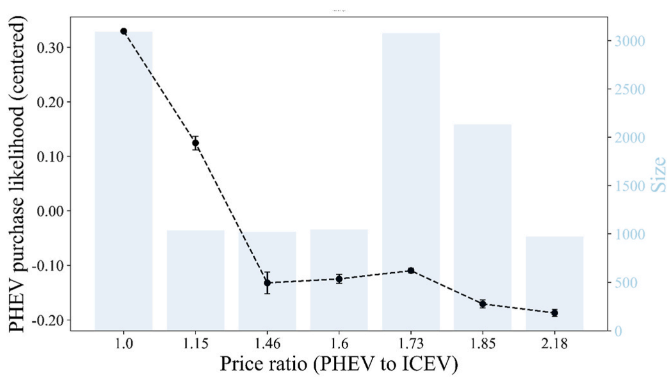


Fig. 7. The influence of PHEV to ICEV price ratio on the likelihood of buying a PHEV.

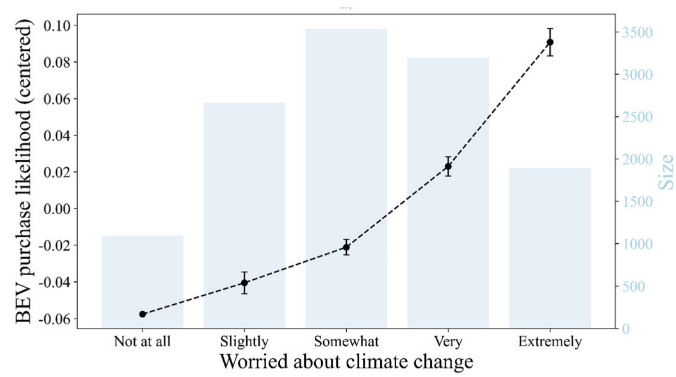


Fig. 8. The influence of worried about climate change on the likelihood of buying a BEV.

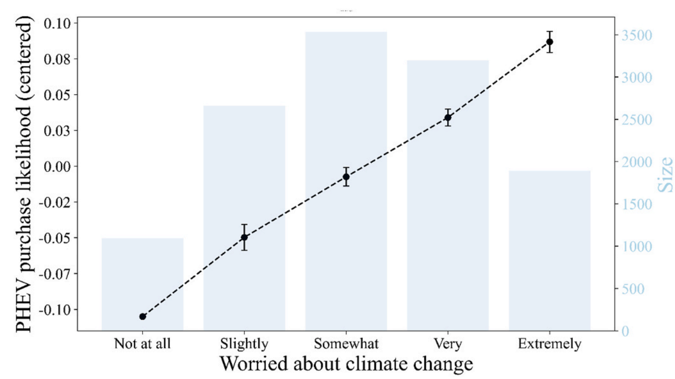


Fig. 9. The influence of worried about climate change on the likelihood of buying a PHEV.

As shown in Fig. 6, increasing the BEV to ICEV price ratio from 1.33 to 2.15, the BEV purchase likelihood is sharply reduced. That is, increasing this ratio from 1.33 to 1.54, 1.60, and 2.15 reduces the BEV purchase likelihood by approximately 12%, 14%, and 18%, respectively. If the BEV to ICEV price ratio increases from 2.15 to 2.54, the BEV purchase likelihood is not considerably dropped.

PHEV purchase likelihood follows a similar trend for the price ratio: a sharp decrease and then remaining the same. Increasing the PHEV to ICEV price ratio from 1 to 1.15 and 1.46 reduces the PHEV purchase likelihood by 20% and 48%, respectively. Accordingly, changes in PHEV purchase likelihood are significantly higher than that of BEVs by changing the price ratios. Moreover, the intention to buy PHEVs is considerably high when their prices are equal to ICEVs. In the United States, high purchase prices of EVs were found significant barriers to buying EVs (Hidrue et al., 2011). Further, reducing EV prices could increase the sales of EVs in Tenerife (Ramos-Real et al., 2018). Those results are in line with the results of this study, while this study provides further details about the influence of purchase price on the EV purchase likelihood.

4.3.1.2. Worried about climate change. Worried about climate change is the first and second variable with the highest relative influence of the intention to buy BEVs and PHEVs, in the order mentioned. The influence of the level of worried about climate change on the willingness to purchase BEVs and PHEVs is illustrated in Figs. 8 and 9. There is a direct relation between the level of worried about climate change and the likelihood of buying an EV. As indicated in Fig. 8, the BEV purchase likelihood is positive for only those who are very worried or extremely worried. These two groups represent roughly 41 % of our sample. The range of influence goes from −6 % to +8 % for the likelihood to choose a BEV over an ICEV.

Similarly, increasing the level of worried about climate change results in an increment in the PHEV purchase likelihood. Similar to BEVs, only those at the two highest levels of worry are positively inclined to PHEVs. The range of influence is slightly larger, going from −10 % to +9 %. Previous studies showed that environmental concern is an important parameter in the acceptance and uptake of EVs (Rezvani et al., 2015). Therefore, the results of this analysis are line with the results of previous studies.

4.3.1.3. Car ownership. Car ownership is the second and four top variables influencing the intention to buy PHEVs and BEVs. The effect of car ownership on the willingness to purchase BEVs and PHEVs is displayed in Figs. 10 and 11. As shown, the probability of buying a BEV and a

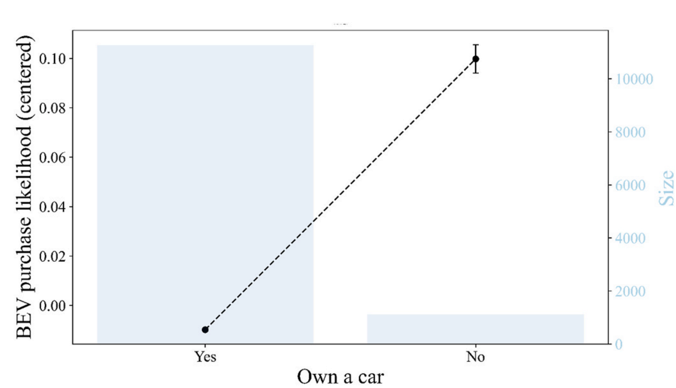


Fig. 10. The influence of car ownership on the likelihood of buying a BEV.

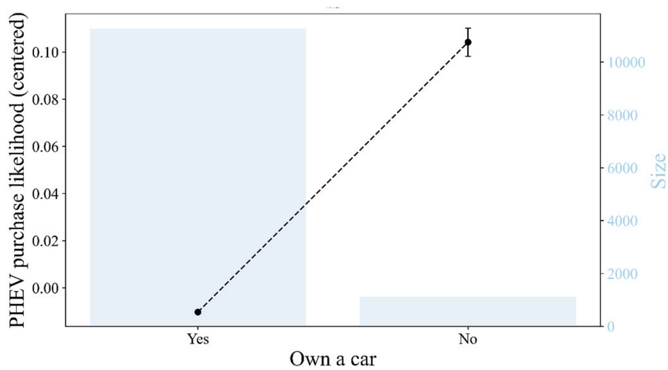


Fig. 11. The influence of car ownership on the likelihood of buying a PHEV.

PHEV among car owners is 11% and 12% less than that of those who do not own a car.

4.3.1.4. CC-SoC. The relative influence of CC-SoC on the intention to buy BEVs and PHEVs is 8%. The influence of CC-SoC on the willingness to purchase BEVs and PHEVs is presented in Figs. 12 and 13. As can be seen, CC-SoC4 (will reduce emissions) and CC-SoC5 (have reduced emissions) groups are inclined to choose BEVs, while the others are not. These two higher stage groups represent roughly 64 % of the sample. The two lowest are at similar negative levels.

For PHEVs, individuals at stage 3 are neutral, with those at the highest two stages more likely to choose a PHEV. The least inclined to choose a PHEV are those who identify as being concerned about climate change but not intending to make any changes to reduce their emissions. The range of difference is larger for PHEVs (roughly −11% to + 4%) than for BEVs (roughly −5% to + 2%). Waygood et al. (2021) investigated the influence of CC-SoC on willingness to pay for EVs in the United States. Their outcomes showed that people at the highest stage of CC-SoC are most likely to pay for EVs, while the CC-SoC1 group is least likely to pay for EVs. The results of their study are consistent with the BEV purchase likelihood. However, the CC-SoC4 group (will reduce emissions) is most likely to choose PHEVs over ICEVs, which is additional details provided by the current study.

4.3.1.5. GHG information framing. The influence of different vehicle labels (i.e., treatments) on BEV and PHEV purchase likelihood is demonstrated in Figs. 14 and 15. As shown, all designed treatments outperform NRCan (the current label in Canada) in terms of increasing the probability of buying BEVs. Leaves framing maximizes the BEV

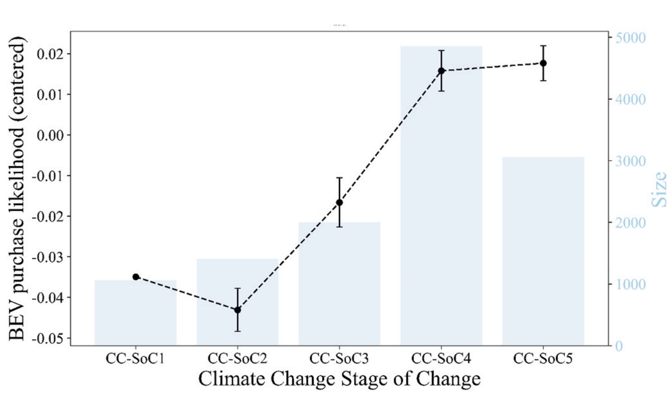


Fig. 12. Influence of CC-SoC on BEV purchase likelihood where: CC-SoC1 = not concerned; CC-SoC2 = concerned, but don't plan to reduce emissions; CC-SoC3 = concerned, but don't know what to do; CC-SoC4 = concerned, and planning to reduce emissions; CC-SoC5 = concerned and have significantly reduced emissions.

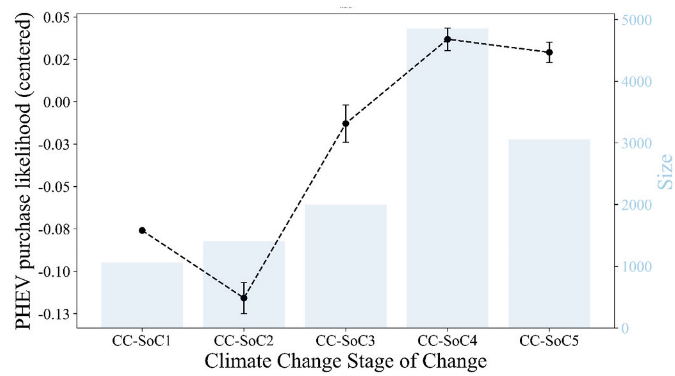


Fig. 13. Influence of CC-SoC on PHEV purchase likelihood where: CC-SoC1 = not concerned; CC-SoC2 = concerned, but don't plan to reduce emissions; CC-SoC3 = concerned, but don't know what to do; CC-SoC4 = concerned, and planning to reduce emissions; CC-SoC5 = concerned and have significantly reduced emissions.

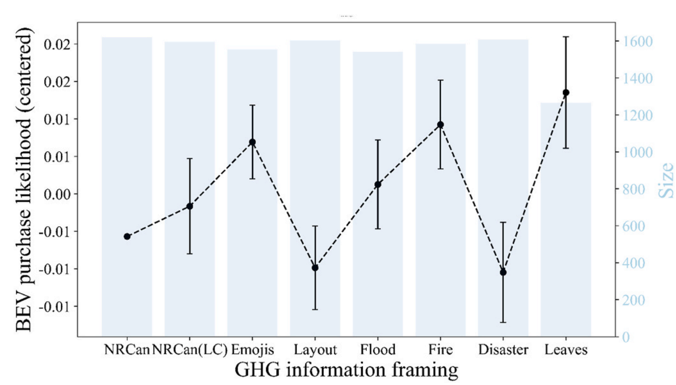


Fig. 14. Influence of GHG information framing on BEV purchase likelihood.

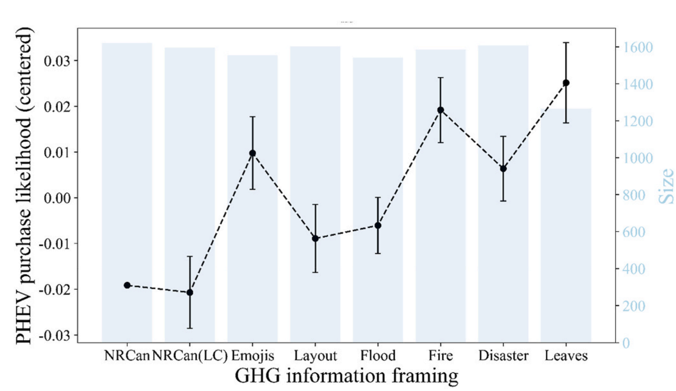


Fig. 15. Influence of GHG information framing on PHEV purchase likelihood.

purchase likelihood, followed by fire, emojis, flood, NRCan with Life Cycle (LC) emissions, NRCan, layout, and disaster. Replacing treatment 8 (leaves) with the current mock-up (NRCan) can increase the probability of buying BEVs by around 3%.

For PHEVs, the leaves framing is the most effective label, followed by fire, emojis, disaster, flood, layout, NRCan, and NRCan with LC emissions. In this case, NRCan is more effective than NRCan with LC emissions, because the GHG emissions of PHEVs are more when only tailpipe emissions are taken into account. Using leaves instead of the current labels in Canada, the likelihood of buying PHEVs increases by approximately 4.5%. The emojis framing was the most effective label in a previous study (Xun et al., 2022). However, leaves and fire (designed in this study) are found to be more effective than emojis in terms of attracting individuals to choose EVs over ICEVs.

4.3.2. ALE results for variable interactions

The interaction between worried about climate change and purchase price ratio has the highest relative influence on the willingness to buy BEVs. Moreover, the interaction between CC-SoC and purchase price ratio has the highest contribution to the intention to buy PHEVs. Therefore, these two interactions are investigated using ALE, and the results are presented in this section.

The influence of the interaction between worries about climate change and the BEV to ICEV purchase price ratio is indicated in Fig. 16. As can be seen, those extremely worried about climate change are less likely to buy cheaper BEVs since their GHG savings are less than the more expensive BEV options. That is, those who are extremely worried about climate change are more likely to prefer BEVs when their GHG reduction is higher while they are more expensive. On the other hand, individuals who are very concerned about climate change are more likely to prefer BEVs over ICEVs when the purchase price ratio is around 1.8. Those who are not worried or slightly worried about climate change are more likely to buy a BEV when the purchase price of BEV to ICEV is the minimum level.

The influence of the interaction between CC-SoC and PHEV to ICEV purchase price ratio is shown in Fig. 17. People who are not concerned about climate change (CC-SoC1) and who concerned, but don't know what to do (CC-SoC2) are more likely to buy PHEVs when the purchase price of PHEVs equals to ICEVs. That is, they are not willing to pay additional money to reduce their emissions. On the other hand, people at the highest stages of change (CC-SoC5 and CC-SoC4) prefer PHEVs over ICEVs when the GHG saving is high, and as a result, they choose more expensive PHEV options with higher GHG reduction.

5. Conclusions

The aim of this study was to generate accurate models to predict who prefers battery electric vehicles and plug-in hybrid electric vehicles over internal combustion engine vehicles. Moreover, this study attempts to determine which variables influence the intention to buy electric vehicles, and what is the influence direction of top variables on the purchase likelihood of electric vehicles. Hence, a DCE is designed, and 2062 participants join the survey. Each participant should select between BEVs and ICEVs in six choice tasks and choose between PHEVs and ICEVs in six other choice tasks. Different vehicle labels are designed to find the optimal GHG information framing.

Three ensemble learning techniques are applied for the modeling: XGB, LGBM, and CatBoost. LGBM outperforms other techniques for predicting individual choices between BEVs and ICEVs, with a testing data accuracy of 82.7%. However, XGB is the most accurate method for predicting participant choices between PHEVs and ICEVs, and it reaches

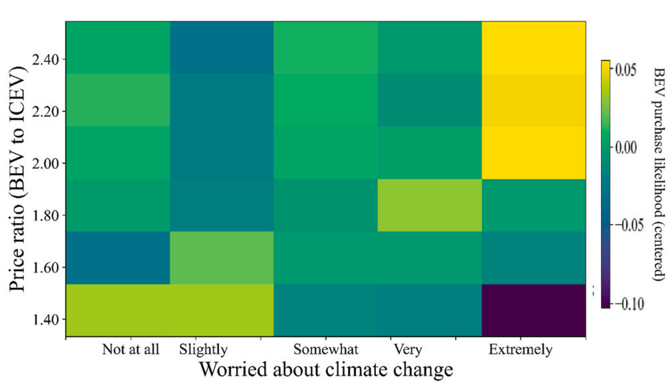


Fig. 16. Influence of the interaction between BEV to ICEV price ratio and worried about climate change on BEV purchase likelihood.

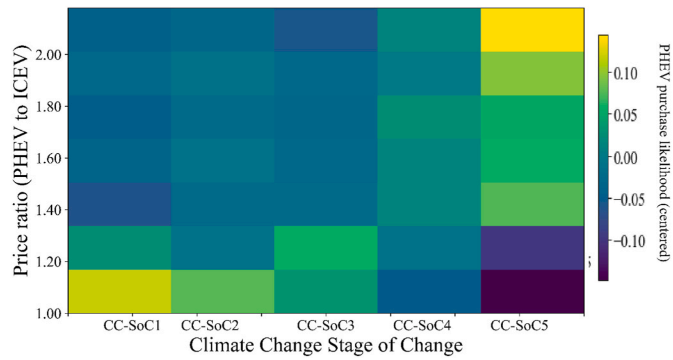


Fig. 17. Influence of the interaction between PHEV to ICEV price ratio and CC-SoC on BEV purchase likelihood.

a prediction accuracy of 79.7%. The results suggest that the level of worried about climate change, car ownership, and purchase price ratio of BEV to ICEV are the most influential parameters on the intention to buy BEVs. Further, PHEV to ICEV purchase price ratio, level of worry about climate change, and climate change stage of change are the three variables with the highest relative influence on the willingness to buy PHEVs. In terms of policy, the other variable of interest was the framing of GHG emissions, which had a relative influence of 5.4 % for both EV types.

Then, the xgbfir library was employed to investigate the interaction of variables on the intention to buy BEVs and PHEVs. The interaction between the purchase price ratio and the level of worry about climate change is the most influential variable interaction on the willingness to buy BEVs. The interaction between the purchase price ratio and the climate change stage of change has the highest relative influence on the intention to buy PHEVs.

Since machine learning techniques are black-box, they cannot illustrate the influence direction of variables on the response variable. To overcome this issue, Accumulated Local Effect (ALE) was used to investigate the influence direction of top variables and variable interactions on the intention to buy EVs. ALE outcomes reveal that those who are extremely worried about climate change, do not own a car, and self-identified as being at the top of the CC-SoC (CC-SoC4 & 5) are more likely to buy electric vehicles. Moreover, increasing the purchase price of electric vehicles can reduce the probability of buying electric vehicles sharply with 1.6 being the threshold for BEVs and roughly 1.3 for PHEVs.

Regarding the GHG information framing, the leaves framing is the most efficient label to increase the likelihood of EV purchase, and it can increase the probability of selecting BEVs over ICEVs by 3% compared to the current labels in Canada (NRCan). Similarly, the leaves framing is the most effective label for PHEVs. Applying the optimal label (leaves) rather than the current mock-up can increase the PHEV purchase likelihood by nearly 4.5%. Most of the new framing styles outperformed the NRCan label format. Accordingly, this research reveals further evidence that how presenting GHG information framing is essential in choice.

One of the limitations of this study is to only apply data analysis to investigate EV preferences. It is recommended that empirical methods are also applied in future studies. Another limitation of this study is to only apply online surveys to collect data. Hence, it is recommended that offline surveys are also used for data collection and their results are compared with online surveys.

CRediT authorship contribution statement

Hamed Naseri: Writing – original draft, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **E.O.D. Waygood:** Writing – review & editing, Supervision, Investigation, Data curation, Conceptualization. **Zachary Patterson:** Writing – review &

editing, Supervision, Investigation. **Bobin Wang:** Writing – review & editing, Data curation.

Data availability

The data that has been used is confidential.

Acknowledgments

This research was supported by (a): the Fonds de recherches du Québec – Nature et Technologie (FRQNT) [grant number 2019-GS-261583]; (b): the Fonds de recherches du Québec – Nature et Technologie (FRQNT) [grant number 323214]; (c): Social Sciences and Humanities Research Council [grant number 435-20200-1292] (d): Fonds de recherches du Québec – Nature et Technologie (FRQNT) [grant number 322727]; (e): Centre interdisciplinaire de recherche en operationalisation du développement durable (CIRODD)'s program SYNERGIE."

References

Abotalebi, E., Scott, D.M., Ferguson, M.R., 2019. Why is electric vehicle uptake low in Atlantic Canada? A comparison to leading adoption provinces. J. Transp. Geogr. 74, 289–298. https://doi.org/10.1016/j.jtrangeo.2018.12.001.

Ambrose, H., Kendall, A., Lozano, M., Wachche, S., Fulton, L., 2020. Trends in life cycle greenhouse gas emissions of future light duty electric vehicles. Transport. Res. Transport Environ. https://doi.org/10.1016/j.trd.2020.102287.

Apley, D.W., Zhu, J., 2020. Visualizing the effects of predictor variables in black box supervised learning models. J. R. Stat. Soc. Ser. B Stat. Methodol. 82, 1059–1086. https://doi.org/10.1111/rssb.12377.

Austmann, L.M., 2021. Drivers of the electric vehicle market: a systematic literature review of empirical studies. Finance Res. Lett. 41, 101846 https://doi.org/10.1016/j.frl.2020.101846.

Barros, V., Cruz, C.O., Júdice, T., Sarmento, J.M., 2021. Is taxation being effectively used to promote public transport in Europe? Transport Pol. 114, 215–224. https://doi.org/10.1016/j.tranpol.2021.10.003.

Bas, J., Cirillo, C., Cherchi, E., 2021a. Classification of potential electric vehicle purchasers: a machine learning approach. Technol. Forecast. Soc. Change 168, 120759. https://doi.org/10.1016/j.techfore.2021.120759.

Bas, J., Zou, Z., Cirillo, C., 2021b. An interpretable machine learning approach to understanding the impacts of attitudinal and ridesourcing factors on electric vehicle adoption. Transp. Lett. https://doi.org/10.1080/19427867.2021.2009098.

Boeri, M., Saure, D., Schacht, A., Riedl, E., Hauber, B., 2020. Modeling heterogeneity in Patients' preferences for psoriasis treatments in a multicountry study: a comparison between random-parameters logit and latent class approaches. Pharmacoeconomics 38, 593–606. https://doi.org/10.1007/s40273-020-00894-7.

Brase, G.L., 2019. What would it take to get you into an electric car? Consumer perceptions and decision making about electric vehicles. J. Psychol. Interdiscip. Appl. 153, 214–236. https://doi.org/10.1080/00223980.2018.1511515.

Cai, W., Wei, R., Xu, L., Ding, X., 2022. A method for modelling greenhouse temperature using gradient boost decision tree. Inf. Process. Agric. 9, 343–354. https://doi.org/10.1016/j.inpa.2021.08.004.

Cerwick, D.M., Gkritza, K., Shaheed, M.S., Hans, Z., 2014. A comparison of the mixed logit and latent class methods for crash severity analysis. Anal. Methods Accid. Res. 3–4, 11–27. https://doi.org/10.1016/j.amar.2014.09.002.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, NY, USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Colgan, J.D., Hinthorn, M., 2021. Is cheap gasoline killing us? Fuel subsidies and under-taxation as a driver of obesity and public health problems worldwide. Energy Res. Social Sci. 82, 102316 https://doi.org/10.1016/j.erss.2021.102316.

Daziano, R.A., Waygood, E.O.D., Patterson, Z., Braun Kohlová, M., 2017. Increasing the influence of CO2 emissions information on car purchase. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2017.07.001.

De Souza, L.L.P., Lora, E.E.S., Palacio, J.C.E., Rocha, M.H., Renó, M.L.G., Venturini, O.J., 2018. Comparative environmental life cycle assessment of conventional vehicles with different fuel options, plug-in hybrid and electric vehicles for a sustainable transportation system in Brazil. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2018.08.236.

Dhananjay, B., Sivaraman, J., 2021. Analysis and classification of heart rate using CatBoost feature ranking model. Biomed. Signal Process Control 68, 102610. https://doi.org/10.1016/j.bspc.2021.102610.

Dong, Y., 2022. Analysis of consumers' willingness to accept of government subsidies for electric vehicles. Transport. Res. Procedia 61, 90–97. https://doi.org/10.1016/j.trpro.2022.01.016.

Dong, Y., Sun, Y., Waygood, O., Wang, B., Huang, P., Naseri, H., 2022. Insight into the nonlinear effect of COVID-19 on well-being in China: commuting, a vital ingredient. J. Transport Health 25, 101424. https://doi.org/10.1016/j.jth.2022.101424.

Ferguson, M., Mohamed, M., Higgins, C.D., Abotalebi, E., Kanaroglou, P., 2018. How open are Canadian households to electric vehicles? A national latent class choice analysis with willingness-to-pay and metropolitan characterization. Transport. Res. Transport Environ. https://doi.org/10.1016/j.trd.2017.12.006.

Fevang, E., Figenbaum, E., Fridstrøm, L., Halse, A.H., Hauge, K.E., Johansen, B.G., Raaum, O., 2021. Who goes electric? The anatomy of electric car ownership in Norway. Transport. Res. Transport Environ. 92, 102727 https://doi.org/10.1016/j.trd.2021.102727.

Funke, S.Á., Sprei, F., Gnann, T., Plötz, P., 2019. How much charging infrastructure do electric vehicles need? A review of the evidence and international comparison. Transport. Res. Transport Environ. 77, 224–242. https://doi.org/10.1016/j.trd.2019.10.024.

Government of Canada, 2021. Canada's top 10 weather stories of 2021 - Canada.ca [WWW Document]. URL. https://www.canada.ca/en/environment-climate-change/services/top-ten-weather-stories/2021.html, 11.24.22.

Greene, W.H., Hensher, D.A., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. Transp. Res. Part B Methodol. 37, 681–698. https://doi.org/10.1016/S0191-2615(02)00046-2.

Guo, X., Gao, Y., Zheng, D., Ning, Y., Zhao, Q., 2020. Study on short-term photovoltaic power prediction model based on the Stacking ensemble learning. Energy Rep. 6, 1424–1431. https://doi.org/10.1016/j.egyr.2020.11.006.

Hardman, S., 2019. Understanding the impact of reoccurring and non-financial incentives on plug-in electric vehicle adoption – a review. Transport. Res. Part A Policy Pract. 119, 1–14. https://doi.org/10.1016/j.tra.2018.11.002.

He, H., Wang, C., Wang, S., Ma, F., Sun, Q., Zhao, X., 2021. Does environmental concern promote EV sales? Duopoly pricing analysis considering consumer heterogeneity. Transport. Res. Transport Environ. 91, 102695 https://doi.org/10.1016/j.trd.2021.102695.

Hidrue, M.K., Parsons, G.R., Kempton, W., Gardner, M.P., 2011. Willingness to pay for electric vehicles and their attributes. Resour. Energy Econ. 33, 686–705. https://doi.org/10.1016/j.reseneeco.2011.02.002.

Hillel, T., Bierlaire, M., Elshafie, M., Jin, Y., 2019. Weak teachers: assisted specification of discrete choice models using ensemble learning. Hear. 2019, 8th Symp. Eur. Assoc. Res. Transp. 1–12.

Hussain, S., Mustafa, M.W., Jumani, T.A., Baloch, S.K., Alotaibi, H., Khan, I., Khan, A., 2021. A novel feature engineered-CatBoost-based supervised machine learning framework for electricity theft detection. Energy Rep. 7, 4425–4436. https://doi.org/10.1016/j.egyr.2021.07.008.

Ibrahim, A.A., Ridwan, R.L., Muhammed, M.M., Abdulaziz, R.O., Saheed, G.A., 2020. Comparison of the CatBoost classifier with other machine learning methods. Int. J. Adv. Comput. Sci. Appl. 11, 738–748. https://doi.org/10.14569/IJACSA.2020.0111190.

Jahanbakhsh, H., Karimi, M.M., Naseri, H., Nejad, F.M., 2020. Sustainable asphalt concrete containing high reclaimed asphalt pavements and recycling agents: performance assessment, cost analysis, and environmental impact. J. Clean. Prod. 244, 118837 https://doi.org/10.1016/j.jclepro.2019.118837.

Jamal, A., Zahid, M., Tauhidur Rahman, M., Al-Ahmadi, H.M., Almoshaogeh, M., Farooq, D., Ahmad, M., 2021. Injury severity prediction of traffic crashes with ensemble machine learning techniques: a comparative study. Int. J. Inj. Control Saf. Promot. 28, 408–427. https://doi.org/10.1080/17457300.2021.1928233.

Jeon, H., Seo, W., Park, E., Choi, S., 2020. Hybrid machine learning approach for popularity prediction of newly released contents of online video streaming services. Technol. Forecast. Soc. Change 161, 120303. https://doi.org/10.1016/j.techfore.2020.120303.

Kim, E.J., 2021. Analysis of travel mode choice in seoul using an interpretable machine learning approach. J. Adv. Transport. https://doi.org/10.1155/2021/6685004.

Kim, K., Yang, H., Yi, J., Son, H.E., Ryu, J.Y., Kim, Y.C., Jeong, J.C., Chin, H.J., Na, K.Y., Chae, D.W., Han, S.S., Kim, S., 2021. Real-time clinical decision support based on recurrent neural networks for in-hospital acute kidney injury: external validation and model interpretation. J. Med. Internet Res. 23, e24120 https://doi.org/10.2196/24120.

Kosai, S., Zakaria, S., Che, H.S., Hasanuzzaman, M., Rahim, N.A., Tan, C., Ahmad, R.D.R., Abbas, A.R., Nakano, K., Yamasue, E., Woon, W.K., Amer, A.H.A., 2022. Estimation of greenhouse gas emissions of petrol, biodiesel and battery electric vehicles in Malaysia based on life cycle approach. Sustain. Times 14, 5783. https://doi.org/10.3390/su14105783.

Krishnan, V.V., Koshy, B.I., 2021. Evaluating the factors influencing purchase intention of electric vehicles in households owning conventional vehicles. Case Stud. Transp. Policy 9, 1122–1129. https://doi.org/10.1016/j.cstp.2021.05.013.

Kucukoglu, I., Dewil, R., Cattrysse, D., 2021. The electric vehicle routing problem and its variations: a literature review. Comput. Ind. Eng. 161, 107650 https://doi.org/10.1016/j.cie.2021.107650.

Lamb, W.F., Grubb, M., Diluiso, F., Minx, J.C., 2022. Countries with sustained greenhouse gas emissions reductions: an analysis of trends and progress by sector. Clim. Pol. 22, 1–17. https://doi.org/10.1080/14693062.2021.1990831.

Lashari, Z.A., Ko, J., Jang, J., 2021. Consumers' intention to purchase electric vehicles: influences of user attitude and perception. Sustain. Times 13, 6778. https://doi.org/10.3390/su13126778.

Li, X., Clark, C.D., Jensen, K.L., Yen, S.T., English, B.C., 2013. Consumer purchase intentions for flexible-fuel and hybrid-electric vehicles. Transport. Res. Transport Environ. 18, 9–15. https://doi.org/10.1016/j.trd.2012.08.001.

Li, Z., Ci, Y., Chen, C., Zhang, G., Wu, Q., Qian, Z., Sean, Prevedouros, P.D., Ma, D.T., 2019. Investigation of driver injury severities in rural single-vehicle crashes under rain conditions using mixed logit and latent class models. Accid. Anal. Prev. 124, 219–229. https://doi.org/10.1016/j.aap.2018.12.020.

Lim, S., Dolsak, N., Prakash, A., Tanaka, S., 2022. Distributional concerns and public opinion: EV subsidies in the U.S. and Japan. Energy Pol. 164, 112883 https://doi.org/10.1016/j.enpol.2022.112883.

Ling, Z., Cherry, C.R., Wen, Y., 2021. Determining the factors that influence electric vehicle adoption: a stated preference survey study in Beijing, China. Sustain. Times 13, 11719. https://doi.org/10.3390/su132111719.

Liu, Y., Lyu, C., Khadka, A., Zhang, W., Liu, Z., 2020. Spatio-temporal ensemble method for car-hailing demand prediction. IEEE Trans. Intell. Transport. Syst. 21, 5328–5333. https://doi.org/10.1109/TITS.2019.2948790.

Long, Z., Axsen, J., 2022. Who will use new mobility technologies? Exploring demand for shared, electric, and automated vehicles in three Canadian metropolitan regions. Energy Res. Social Sci. 88, 102506 https://doi.org/10.1016/j.erss.2022.102506.

Louviere, J.J., Hensher, D.A., Swait, J.D., Adamowicz, W., 2000. Stated Choice Methods: Analysis and Applications. Cambridge University Press. Cambridge University Press. https://doi.org/10.1017/cbo9780511753831.

McDonald, R.I., Chai, H.Y., Newell, B.R., 2015. Personal experience and the "psychological distance" of climate change: an integrative review. J. Environ. Psychol. https://doi.org/10.1016/j.jenvp.2015.10.003.

Naseri, H., Golroo, A., Shokoohi, M., Gandomi, A.H., 2022. Sustainable pavement maintenance and rehabilitation planning using the marine predator optimization algorithm. Struct. Infrastruct. Eng. 1–13. https://doi.org/10.1080/15732479.2022.2095407.

Naseri, H., Shokoohi, M., Jahanbakhsh, H., Golroo, A., Gandomi, A.H., 2021a. Evolutionary and swarm intelligence algorithms on pavement maintenance and rehabilitation planning. Int. J. Pavement Eng. 1–15. https://doi.org/10.1080/10298436.2021.1969019.

Naseri, H., Waygood, E.O.D., Wang, B., Patterson, Z., Daziano, R.A., 2021b. A novel feature selection technique to better predict climate change stage of change. Sustainability 14, 40. https://doi.org/10.3390/su14010040.

Nguyen-Sy, T., Wakim, J., To, Q.D., Vu, M.N., Nguyen, T.D., Nguyen, T.T., 2020. Predicting the compressive strength of concrete from its compositions and age using the extreme gradient boosting method. Construct. Build. Mater. 260, 119757 https://doi.org/10.1016/j.conbuildmat.2020.119757.

Niu, D., Diao, L., Zang, Z., Che, H., Zhang, T., Chen, X., 2021. A machine-learning approach combining wavelet packet denoising with catboost for weather forecasting. Atmosphere 12, 1618. https://doi.org/10.3390/atmos12121618.

NRCan, 2019. Oil supply and demand. Nat. Resour. Can. 1, 1–54.

Pham, H., Jiang, X., Zhang, C., 2022. Causality and advanced models in trip mode prediction: interest in choosing swissmetro. OSF Prepr. https://doi.org/10.31219/OSF.IO/M4W38.

Pi, Y., 2021. Machine learning in governments: benefits, challenges and future directions. eJournal eDemocracy Open Gov. 13, 203–219. https://doi.org/10.29379/jedem.v13i1.625.

Pineda-Jaramillo, J., Arbeláez-Arenas, Ó., 2022. Assessing the performance of gradient-boosting models for predicting the travel mode choice using household survey data. J. Urban Plann. Dev. 148 https://doi.org/10.1061/(asce)up.1943-5444.0000830.

Plananska, J., Gamma, K., 2022. Product bundling for accelerating electric vehicle adoption: a mixed-method empirical analysis of Swiss customers. Renew. Sustain. Energy Rev. 154, 111760 https://doi.org/10.1016/j.rser.2021.111760.

Prabhavathi, C., Mahesh, A., Ajay, S.V., Mythili, R., 2022. Malware prediction using XGBOOST and CATBOOST. J. Eng. Sci. 13, 620–626.

Ramos-Real, F.J., Ramírez-Díaz, A., Marrero, G.A., Perez, Y., 2018. Willingness to pay for electric vehicles in island regions: the case of Tenerife (Canary Islands). Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2018.09.014.

Rezvani, Z., Jansson, J., Bodin, J., 2015. Advances in consumer electric vehicle adoption research: a review and research agenda. Transport. Res. Transport Environ. 34, 122–136. https://doi.org/10.1016/j.trd.2014.10.010.

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, 116253 https://doi.org/10.1016/j.eswa.2021.116253.

Shangguan, Q., Fu, T., Wang, J., Fang, S., Fu, L., 2022. A proactive lane-changing risk prediction framework considering driving intention recognition and different lane-changing patterns. Accid. Anal. Prev. 164, 106500 https://doi.org/10.1016/j.aap.2021.106500.

Shen, J., 2009. Latent class model or mixed logit model? A comparison by transport mode choice data. Appl. Econ. 41, 2915–2924. https://doi.org/10.1080/00036840801964633.

Simsekoglu, Ö., Nayum, A., 2019. Predictors of intention to buy a battery electric vehicle among conventional car drivers. Transport. Res. F Traffic Psychol. Behav. 60, 1–10. https://doi.org/10.1016/j.trf.2018.10.001.

Sobiech-Grabka, K., Stankowska, A., Jerzak, K., 2022. Determinants of electric cars purchase intention in Poland: personal attitudes v. Economic arguments. Energies 15, 3078. https://doi.org/10.3390/en15093078.

Tutica, L., Vineel, K., Mishra, S., Mishra, M.K., Suman, S., 2021. Invoice deduction classification using LGBM prediction model. In: Lecture Notes in Electrical Engineering. Springer Science and Business Media Deutschland GmbH, pp. 127–137. https://doi.org/10.1007/978-981-15-8752-8_13.

Vafaei-Zadeh, A., Wong, T.K., Hanifah, H., Teoh, A.P., Nawaser, K., 2022. Modelling electric vehicle purchase intention among generation Y consumers in Malaysia. Res. Transp. Bus. Manag. 43, 100784 https://doi.org/10.1016/j.rtbm.2022.100784.

Wang, B., Waygood, E.O.D., Daziano, R.A., Patterson, Z., Feinberg, M., 2021. Does hedonic framing improve people's willingness-to-pay for vehicle greenhouse gas emissions? Transport. Res. Transport Environ. 98, 102973 https://doi.org/10.1016/j.trd.2021.102973.

Wang, S., Mo, B., Hess, S., Zhao, J., 2020. Comparing hundreds of machine learning classifiers and discrete choice models in predicting travel behavior: an empirical benchmark. arXiv Prepr. arXiv2102, 01130.

Wang, Y., Sherry Ni, X., 2019. A XGBOOST risk model via feature selection and bayesian hyper-parameter optimization. Int. J. Database Manag. Syst. 11, 1–17. https://doi.org/10.5121/ijdms.2019.11101.

Waygood, E.O.D., Wang, B., Daziano, R.A., Patterson, Z., Kohlová, M.B., 2021. The climate change stage of change measure: vehicle choice experiment. J. Environ. Plann. Manag. https://doi.org/10.1080/09640568.2021.1913107.

Xun, J., Waygood, E.O.D., Wang, B., Naseri, H., Loiselle, L.A., Patterson, Zachary, Feinberg, M., 2022. Exploring the effects of new framing techniques for greenhouse gas emissions. In: Transportation Research Board 101st Annual Meeting. Washington DC, United States.

Zhou, Z., Wang, M., Huang, J., Lin, S., Lv, Z., 2021. Blockchain in big data security for intelligent transportation with 6G. IEEE Trans. Intell. Transport. Syst. https://doi.org/10.1109/TITS.2021.3107011.

Zhu, S., Zhu, F., 2019. Cycling comfort evaluation with instrumented probe bicycle. Transport. Res. Part A Policy Pract. 129, 217–231. https://doi.org/10.1016/j.tra.2019.08.009.