Civil Engineering and Applied Solutions

Research Article

journal homepage: ceas.journals.umz.ac.ir

CEAS Civil Engineering and Applied Solutions

A Comparison of Machine Learning Models in Predicting Competition Between High-Speed Rail (HSR) and Air Transport: The Tehran-Mashhad Case Study

Mohammad Feli [a], Ali Naderan [a*], Mahmoud Saffarzadeh [b]

[a] *Department of Civil Engineering, SR.C., Islamic Azad University, Tehran, Iran*
[b] *Department of Civil and Environmental Engineering, Tarbiat Modares University, Tehran, Iran*

ARTICLE INFO

Article history:

Received 26 March 2026

Revised 23 April 2026

Accepted 18 May 2026

Available online 01 January 2027

Keywords:

High-speed rail

Mode choice

Machine learning

Classification report

ABSTRACT

Mode choice modeling represents a fundamental domain within transportation engineering and urban-regional planning. The competition between high-speed rail (HSR) and air travel holds particular significance for demand forecasting, revenue optimization, environmental policy, and infrastructure development. Traditional discrete choice models have long served as the cornerstone of mode choice analysis. These models offer interpretability, compatibility with stated and revealed preference data, and the capacity to compute policy-relevant elasticities. However, they suffer from critical limitations, such as the independence of irrelevant alternatives (IIA) assumption and inability to accommodate large, noisy datasets. Conversely, Machine Learning (ML) methodologies have gained prominence for their capacity to handle complex, nonlinear, and high-dimensional data. By applying Artificial Neural Network (ANN), Random Forest (RF), Decision Tree (DT), and K-Nearest Neighbor (KNN), this study addresses two critical research gaps: (1) the scarcity of ML applications in developing countries with limited HSR infrastructure, and (2) the limited incorporation of psychological factors alongside socio-economic variables. Using stated preference data from 100 Iranian respondents across 18 travel scenarios, this research develops ML-based models, examining variables such as travel time, cost, income, service frequency, previous travel experiences, and psychological factors including fear of flying. The findings reveal that ANN emerged as the top performer with an overall accuracy of 84.67%. The RF model followed with 82.44% accuracy, showing robust predictive capability with relatively balanced class-wise performance, though slightly favoring the majority class. Also, class-specific analysis across all models consistently demonstrated higher precision for airplane predictions.

How to cite this article: Feli, M., Naderan, A. Saffarzadeh, M. A Comparison of Machine Learning Models in Predicting Competition Between High-Speed Rail (HSR) and Air Transport: The Tehran- Mashhad Case Study. Civil Engineering and Applied Solutions. 2027; 3(1): 146–160. doi:10.22080/ceas.2026.31482.1086.

1. Introduction

Mode choice modeling is considered one of the most important and fundamental fields of transportation engineering and urban-regional planning. This field specifically addresses the examination, analysis, and prediction of passengers' or transport users' choice behavior among various transport mode options, where individuals or groups decide which vehicle or combination of vehicles to use for a particular trip. Among the various options, the competition between high-speed rail (HSR) and airplane is one of the most practical topics, as these two modes often compete directly with each other over medium to long distances (typically 300 to 1200 kilometers), and the choice of one has a significant impact on the other, particularly in the areas of demand, operators' revenue,

* Corresponding author.
E-mail addresses: naderan@iau.ac.ir (A. Naderan).

https://doi.org/10.22080/ceas.2026.31482.1086
ISSN: 3092-7749/

environmental policies, and infrastructure development [1, 2].

Mode choice models help predict future demand with greater accuracy, optimize infrastructure, and strengthen sustainable policies. Classic mode choice models, such as the Multinomial Logit (MNL) model, binary logit model, mixed logit models, or Bayesian logit models, have been among the most widely used approaches for decades. These models are built upon discrete choice theory and the assumption of random utility maximization, and they offer important advantages. High interpretability, compatibility with stated preference (SP, i.e., hypothetical scenarios) and revealed preference (RP, i.e., actual observed behavior) data, and the ability to calculate policy-relevant indicators such as demand elasticities are among the key strengths of this modeling approach [3, 4]. Despite the aforementioned positive aspects, this approach also has serious limitations, including the assumption of independence of irrelevant alternatives (IIA), which is often violated; linear or simplified relationships between variables; inability to handle unobserved heterogeneity; and weaker performance in accurate prediction when the data are large and noisy [5].

In recent years, this field has witnessed a remarkable surge in the use of machine learning methods, the main reason for this shift being the high capability of ML in handling large, complex, and nonlinear data; data that include numerous variables such as travel time, cost, income, trip purpose, accessibility, service frequency, demographic characteristics, and even ticket purchasing behavior, which often exhibit nonlinear, interactive, and heterogeneous relationships. The succeeding section will delve into the available literature review addressing mode choice in HSR-airplane competition.

The remainder of this paper is structured as follows. Section 2 reviews the relevant literature on HSR-airplane competition and the evolution of mode choice modeling approaches. Section 3 describes the materials and methods, including data collection, variable specification, and the four machine learning algorithms employed (ANN, RF, DT, and KNN). Section 4 presents the modeling results, including classification reports and confusion matrices for each model. Section 5 provides a comparative discussion of model performance. Finally, Section 6 concludes the paper with key findings, policy implications, and directions for future research.

2. Literature review

The modeling of travel mode choice has been extensively studied through two primary methodological streams: traditional discrete choice models grounded in random utility theory and, more recently, machine learning (ML) approaches. Comprehensive systematic reviews by Ali [6], Fale et al. [7] highlight that while the multinomial logit model remains the most widely used framework due to its interpretability and solid theoretical foundations, it suffers from critical limitations such as the Independence of Irrelevant Alternatives (IIA) assumption, reliance on linear or oversimplified variable relationships, and an inability to fully capture unobserved heterogeneity across decision-makers. The concept of unobserved heterogeneity, referring to variations in preferences or decision processes that are not captured by observed variables, has been a central concern in choice modeling. As noted by Forsythe et al. [8], traditional approaches to modeling heterogeneity, such as mixed logit models, rely on strong parametric assumptions about the distribution of random parameters. In contrast, their proposed Mixed Aggregate Preference Logit (MAPL) model leverages machine learning to relax these assumptions and capture unobserved heterogeneity more flexibly.

The overall trend in the literature review indicates a gradual transition from traditional logit models to advanced machine learning algorithms, which have increased prediction accuracy and enabled the discovery of nonlinear relationships and better interpretability. The evolutionary trend in the literature review has moved from simple comparisons of machine learning with logit to the use of real large-scale data, hybrid frameworks, and advanced models.

To identify the variables affecting this competition, numerous studies have been conducted. Variables such as cost [9, 10], fleet speed [11, 12], travel time and its components [13, 14], individuals' socio-economic characteristics such as age [15], income [16], gender [17], education [15], pollutant and pollution considerations [18, 19], service frequency [20, 21], and comfort and convenience [22, 23]. This research, relying on these variables, creates various functional forms of them, and variables such as previous travel experiences with the train or fear of flying are also considered in the modeling process.

From the perspective of modeling and the approach used, initial studies focused on comparing the performance of machine learning with logit models in intercity mode choice (including HSR, airplane, conventional train, and bus). For example, Li and colleagues compared Bayesian Mixed Multinomial Logit (BMNL), Multinomial Logit (MNL), Multi-Layer Perceptron (MLP), and Radial Basis Function (RBF) neural networks. The results showed that neural networks (especially with balanced data) have higher prediction accuracy, but Bayesian models were superior in handling unobserved heterogeneity. Key factors included travel time, cost, and demographic characteristics [3, 24].

Focus on the application of machine learning in predicting demand and mode choice behavior in specific contexts, such as the impact of the COVID-19 pandemic on HSR in the United States, represents another aspect of the conducted research. Pan's study, in the Los Angeles-San Francisco corridor, using stated preference data, identified factors such as total travel time, accessibility, comfort, travel frequency, gender, and mobility issues. Multinomial Logit Regression models showed accuracy of approximately 67-80 percent, but it was suggested that machine learning could improve accuracy [25].

In a study in Thailand (with a focus on HSR development), the Binary Logit model was compared with machine learning algorithms such as boosting XGBoost, LightGBM, and CatBoost. CatBoost showed the best performance (accuracy ≈0.885, AUC ≈0.958). Influential factors included travel time, cost, household income, service frequency, access time, and trip purpose. The results indicated that XGBoost and CatBoost interpret nonlinear relationships well, and machine learning is generally superior to logit [1].

In a study using real large-scale ticketing data and ticket sales integration, the MNL model was compared with eight algorithms including RF, SVM, and KNN, and the results showed that RF had the best prediction accuracy. This study proposed a new framework for mode choice analysis using real data [2].

Extending their previous work in Thailand, Banyong et al. (2025) used large-scale stated preference data (3,200 respondents) to compare the MNL model with XGBoost, LightGBM, CatBoost, Deep Neural Network (DNN), and Convolutional Neural Network (CNN). The results showed that CatBoost was the superior model (AUC ≈ 0.911, accuracy ≈ 0.756). Key factors included travel cost, service frequency, and waiting time. Tree-based models such as CatBoost were more stable and resistant to overfitting in large datasets compared to deep neural networks. This study emphasized the application of machine learning in developing countries for HSR policy-making [4].

While the reviewed studies demonstrate the superiority of machine learning over traditional logit models in capturing nonlinear relationships and improving prediction accuracy, there remains a paucity of applications in developing countries with limited high-speed rail infrastructure. Moreover, few investigations incorporate psychological factors such as fear of flying or prior travel experiences alongside socio-economic variables.

Although the competition between high-speed rail and air travel has been extensively studied in developed countries, its investigation in developing nations, and particularly in Iran, remains critically limited. The Tehran-Mashhad corridor, as one of the busiest domestic air routes in the Middle East, offers a unique yet underexplored context for mode choice modeling. With a distance of approximately 900 kilometers, high passenger demand, and ongoing discussions about HSR infrastructure development in Iran, understanding the behavioral and environmental factors influencing mode choice in this corridor has direct implications for transportation policy, investment prioritization, and emission reduction strategies. This study, therefore, focuses specifically on the Tehran-Mashhad scenario to provide the first machine learning-based mode choice analysis in Iran that incorporates psychological variables alongside socio-economic attributes.

2.1. Application of artificial neural networks (ANN) in mode choice

Artificial Neural Networks (ANNs) have gained considerable attention in travel mode choice modeling due to their ability to capture complex, nonlinear relationships without imposing predefined utility structures. Unlike traditional discrete choice models, ANNs learn patterns directly from data through interconnected layers of neurons, making them particularly suitable for high-dimensional and noisy datasets. In the context of intercity travel, Li et al. [3] compared Bayesian mixed multinomial logit, multinomial logit, MLP, and RBF neural networks. Their results showed that neural networks, especially when trained on balanced data, achieved higher prediction accuracy than logit models, although Bayesian models remained superior in handling unobserved heterogeneity. Similarly, in the competition between high-speed rail and air travel, ANNs have been employed to model passenger preferences using variables such as travel time, cost, income, and psychological factors [7, 26]. More recent studies have demonstrated that deep neural networks (DNNs) can achieve accuracy levels above 85% in binary mode choice tasks, outperforming traditional logit models in predictive performance [4, 27].

These findings suggest that ANNs are a viable alternative or complement to classical utility-based approaches, particularly when the underlying decision process involves strong nonlinearities and interaction effects. Also, in transportation regression problems, neural networks are employed to predict continuous values such as travel demand, travel time, traffic flow, or energy consumption. For instance, LSTM and DNN models, using inputs such as speed, density, weather, fuel price, and historical data, forecast hourly or daily demand with high accuracy (often above 85–90%) [27, 28]. These models are widely used in Intelligent Transportation Systems (ITS) and navigation applications, although they require large volumes of data and careful hyperparameter tuning [29, 30].

2.2. Application of random forest (RF) in mode choice

Random Forest (RF) is an ensemble learning method that builds multiple decision trees on bootstrapped samples and aggregates their predictions through majority voting (classification) or averaging (regression). In transportation mode choice research, RF has become increasingly popular because of its robustness against overfitting, ability to handle mixed data types (numerical and categorical), and built-in feature importance measures that enhance interpretability. Cao et al. [2] compared the Multinomial Logit model with eight machine learning algorithms, including Random Forest, Support Vector Machine, and K-Nearest Neighbors, using large-scale ticketing data. Their results indicated that Random Forest achieved the best prediction accuracy among all tested models. In the specific context of HSR-airplane competition, RF has been applied to stated preference data incorporating socio-economic variables, trip attributes, and environmental awareness, achieving accuracies of 82-85% [31, 32]. Furthermore, RF's resistance to noise and missing values makes it particularly suitable for developing-country contexts where survey data may contain irregularities [33]. Several studies have also used feature importance from RF to identify key determinants of mode choice, such as travel cost, service frequency, and waiting time [1, 4]. Some features of RFs have made them a practical tool to model regression and classification problems in transportation, such as handling mixed types of data, handling missing data, and outlier robustness [6, 32, 34].

2.3. Application of decision tree (DT) in mode choice

Decision Trees (DT) are among the most interpretable machine learning algorithms, partitioning the feature space recursively based on criteria such as Gini impurity or information gain. Each internal node represents a decision rule on a single feature, and leaf nodes provide the final class prediction. In mode choice modeling, decision trees have been used to generate simple, rule-based representations of traveler behavior, which are highly valuable for policy-making and transport planning [35, 36]. For instance, decision trees can reveal threshold effects, such as "if travel time by HSR exceeds 5 hours and ticket price is above 20 million IRR, then the probability of choosing air travel increases significantly." While individual decision trees are prone to overfitting, they serve as building blocks for more robust ensemble methods like Random Forest and Gradient Boosting. In comparative studies, decision trees typically achieve accuracy levels between 70% and 85% in mode choice classification tasks, depending on tree depth and pruning strategies [37, 38]. Despite their lower predictive power compared to ensemble methods, decision trees remain widely used for exploratory analysis and for generating interpretable if-then rules that can directly inform transport policies [39].

2.4. Application of k-nearest neighbor (KNN) in mode choice

The K-Nearest Neighbors (KNN) algorithm is a non-parametric, instance-based learning method that classifies a new observation based on the majority class among its k closest neighbors in the feature space, using distance metrics such as Euclidean or Manhattan distance. KNN has been applied in mode choice modeling due to its simplicity, absence of a training phase, and ability to adapt to local data structures without assuming any functional form [40, 41]. In the context of intercity travel, KNN has been compared with logit models and other machine learning algorithms, showing competitive performance on moderate-sized datasets. For example, Cao et al. [2] included KNN in their benchmark of eight algorithms and found that KNN achieved reasonable accuracy, though it was outperformed by tree-based ensembles. KNN is particularly effective when the decision boundary between modes is irregular and when the dataset is not excessively large, as computational cost grows with the number of samples and features [34]. However, KNN is sensitive to feature scaling and the choice of k, requiring careful preprocessing. In HSR-airplane choice studies, KNN has achieved accuracy levels of approximately 78-80%, with higher precision typically observed for the more frequent mode (e.g., airplane) [40, 42]. Its transparency and ease of implementation make KNN a useful baseline model for evaluating more complex algorithms.

3. Material and methods

This section describes the data collection process, variable specification, preprocessing steps, model implementation details, and evaluation metrics used in this study.

3.1. Data collection and sample characteristics

An online questionnaire was designed to collect stated preference data from passengers traveling on the Tehran-Mashhad corridor, one of the busiest domestic routes in Iran. The survey was distributed to 100 respondents, each of whom was presented with 18 hypothetical travel scenarios. In each scenario, respondents were asked to choose between High-Speed Rail (HSR) and Airplane (APL) based on a set of trip attributes. The questionnaire also collected socio-economic and psychological characteristics of the respondents.

Stated preference (SP) refers to data collected from hypothetical scenarios where respondents are asked to choose among travel alternatives that are not yet available (e.g., high-speed rail). Revealed preference (RP) refers to data collected from actual past travel decisions, reflecting real-world behavior. This study uses SP data because HSR is not yet operational in the Tehran-Mashhad corridor, making RP data unavailable. SP surveys allow researchers to model choice behavior for new transport services

The socio-economic variables included age (AGE), gender (GENDER), education level (EDU), personal income (INCOME), household income (HHINCOME), previous experience with HSR (HSREXP), number of air trips in the past year (LYAPLTRV), cost of air travel in the past year (LYAPLCST), percentage of monthly income spent on intercity travel (PRCINCMINCTY), maximum acceptable ticket price for Tehran-Mashhad HSR (MAXHSRTCKTM), and affordability for purchasing airplane tickets (ATEFFRD).

Psychological and attitudinal variables included importance of environmental protection in daily decision-making (ENVDDCS), impact of environmental variables on mode choice (ENVEFFCTIMP), awareness of negative impacts of air transportation (ANGTIMPCT), awareness of lower emissions of HSR (LWHSRPLT), willingness to pay more for reduced pollution (MRCSTLWPLT), importance of weekly service frequency (WKFRQ), likelihood of choosing HSR if its frequency exceeds that of APL (HSRMRWFRQ), suitable number of weekly HSR services (HSRSFCWFRQ), fear of flying (APLFEAR), safety concerns during flight (SFTCN), and being forced to use another option due to fear of flying (APLFEALT).

Trip-related attributes varied across the 18 scenarios, including travel time (TIME_HSR, TIME_APL in hours), ticket cost (CST_HSR, CST_APL in Million IRR), availability of comfortable seats (CMSEAT_HSR, CMSEAT_APL), and availability of entertainment (ENTRTMNT_HSR, ENTRTMNT_APL).

3.2. Variable specification

For modeling in this research, the following variables were collected through an online questionnaire from 100 individuals. Passengers with various socio-economic characteristics were presented with 18 scenarios and asked to select their preferred travel mode. The collected variables include those listed in Table 1. In this table, APL stands for airplane, and HSR stands for high-speed rail. In addition to the variables in Table 1, Table 2 also examines the levels of variation in three fundamental variables, namely travel time, cost, and utility. It should be noted that the variables in this research are categorized according to the following description:

- ✓ AGE: age of respondent
- ✓ GENDER: gender of respondent

- ✓ EDU: education level of respondent
- ✓ INCOME: income level of respondent
- ✓ HHINCOME: household income level of respondent
- ✓ HSREXP: previous experience of traveling by HSR
- ✓ LYAPLTRV: the number of air trips in the past year
- ✓ LYAPLCST: the cost of air travel in the past year
- ✓ PRCINCMINCTY: the percentage of monthly income spent on intercity travel
- ✓ MAXHSRTCKTM: the maximum ticket price for the Tehran-Mashhad HSR
- ✓ ATEFFRD: respondents' affordability for purchasing APL tickets
- ✓ ENVDDCS: importance of environmental protection in daily decision-making
- ✓ ENVEFFCTIMP: impact of environmental variables on travel mode choice
- ✓ ANGTIMPCT: awareness of the negative impacts of air transportation
- ✓ LWHSRPLT: awareness of the lower emissions of HSR
- ✓ MRCSTLWPLT: willingness to pay more for reduced pollution
- ✓ WKFRQ: importance of weekly service frequency in mode choice
- ✓ HSRMRWFRQ: likelihood of choosing HSR if its service frequency exceeds that of APL
- ✓ HSRSFCWFRQ: the suitable number of weekly high-speed rail services
- ✓ APLFEAR: fear of flying
- ✓ SFTCN: safety concerns during a flight
- ✓ APLFEALT: being forced to use another option due to fear of flying
- ✓ CST_HSR or CST_APL: Cost of ticket for HSR and APL respectively.
- ✓ TIME_HSR or TIME_APL: In-vehicle travel time in hour for HSR and APL respectively.
- ✓ CMSEAT_HSR or CMSEAT_APL: Availability of comfortable seat for HSR and APL respectively.
- ✓ ENTRTMNT_HSR or ENTRTMNT_APL: Availability of entertainment for HSR and APL respectively.

Table 1. Description of socio-economic feature collected for this study.

Variables					
AGE		**EDU**		**LYAPLTRV**	
Below 20	2%	Associate diploma and lower	3%	No trip	27%
20 to 30	9%	Bachelor	18%	1 to 4 trips	42%
30 to 40	34%	Masters	58%	5 to 8 trips	13%
40 to 50	32%	PhD and equivalent	21%	9 to 16 trips	11%
More than 50	23%	**INCOME (Million IRR)**		More than 16 trips	7%
GENDER		Below 65	2%	**HHINCOME (Million IRR)**	
Male	66%	65 to 115	4%	Less than 200	20%
Female	34%	115 to 150	6%	200 to 350	26%
HSREXP		150 to 175	7%	350 to 500	29%
Yes	29%	175 to 200	13%	500 to 650	19%
No	71%	200 to 230	9%	**ATEFFRD**	
LYAPLCST (Million IRR)		230 to 267	7%	Poor	20%
Below 50	36%	267 to 315	10%	Fairly poor	26%
50 to 100	15%	315 to 390	15%	Average	40%
100 to 150	23%	More than 390	27%	Fairly good	11%
150 to 300	17%	**PRCINCMINCTY**		Good	3%
More than 300	9%	Below 5%	40%	**ENVDDCS**	
MAXHSRTCKTM (Million IRR)		50 to 10 %	37%	Not important	2%

Below 15	43%	10 to 20%	18%	Low	10%
15 to 20	42%	More than 20%	5%	Medium	22%
20 to 30	15%	**ENVEFFCTIMP**		High	28%
More than 30	0%	Not important	2%	Very high	38%
LWHSRPLT		Low	10%	ANGTIMPCT	
Yes	78%	Medium	22%	Yes	60%
No	22%	High	28%	No	40%
MRCSTLWPLT		Very high	38%	**HSRMRWFRQ**	
Not paying	24%	**WKFRQ**		Not important	4%
To 5%	25%	Not important	5%	Low	3%
5 to 10%	26%	Low	8%	Medium	31%
10% to 20%	19%	Medium	42%	High	47%
More than 20%	6%	High	33%	Very high	15%
HSRSFCWFRQ		Very high	12%	**SFTCN**	
Below 5 trips	13%	**APLFEAR**		Not worried	14%
5 to 10 trips	23%	Yes	22%	Low	26%
10 to 15 trips	30%			Medium	24%
15 to 20 trips	16%	No	78%	High	21%
More than 20 trips	18%			Very high	15%

Table 2. Values of cost, time, and utility in scenarios.

Variable	Values
CST_HSR	10, 15, 20 (Million IRR)
TIME_HSR	4, 5, 6 (Hour)
CMSEAT_HSR	"Yes", "No"
ENTRTMNT_HSR	"Yes", "No"
CST_APL	20, 30, 40 (Million IRR)
TIME_APL	1, 1.5, 2 (Hour)
CMSEAT_ APL	"Yes", "No"
ENTRTMNT_ APL	"Yes", "No"

3.3. Model selection rationale

Four machine learning algorithms were selected to cover a diverse range of learning paradigms: ANN represents connectionist models capable of capturing highly nonlinear interactions through multiple layers [4]RF is an ensemble method that reduces overfitting via bagging and provides built-in feature importance [32]DT offers maximum interpretability and transparent decision rules [42]; and K-Nearest Neighbors (KNN) is a simple, non-parametric instance-based learner [31]. These models were chosen because they are widely used in transportation mode choice studies, they differ substantially in their underlying mechanisms (neural, tree-based, distance-based), and they allow a fair comparison between complex (ANN, RF) and simpler (DT, KNN) algorithms. Other powerful methods such as gradient boosting (XGBoost, LightGBM, CatBoost) or deep learning architectures (CNNs, RNNs) are certainly suitable, but they typically require larger datasets or are optimized for specific data structures (e.g., images, sequences). Given our moderate sample size (1,800 observations) and the tabular nature of our data, we focused on these four well-established and computationally efficient models.

3.4. Data preprocessing

Prior to model training, the following preprocessing steps were applied:

Handling missing values: No missing values were present in the collected dataset, as all questionnaire items were mandatory.

Train-test split: The dataset, consisting of 100 respondents × 18 scenarios = 1,800 observations, was randomly split into training (75%, i.e., 1,350 samples) and testing (25%, i.e., 450 samples) sets. The test set contained 200 HSR and 250 APL instances (slight imbalance), as reported in the classification reports. To ensure reproducibility, a fixed random seed (42) was used.

Due to the moderate dataset size, hyperparameters were selected based on default values and existing literature, and the final models were evaluated only on the held-out test set. Future work could incorporate cross-validation for more robust hyperparameter optimization. Also, in future research directions, applying metaheuristic algorithms, as described in [43, 44], is practical and can be considered.

3.5. Evaluation metrics

To evaluate the performance of the four machine learning models, this study uses standard classification metrics commonly employed in imbalanced binary prediction tasks. A brief introduction to these metrics is provided here for readers less familiar with machine learning terminology.

Accuracy: The proportion of all predictions that are correct. While intuitive, accuracy can be misleading when class sizes are unequal.

Precision (for a given class): Among all instances predicted as that class, how many were truly that class. High precision means few false alarms (Type-I errors).

Recall (also known as Sensitivity or True Positive Rate): Among all actual instances of that class, how many were correctly predicted. High recall means few missed detections (Type-II errors).

F1-score: The harmonic mean of precision and recall. It provides a single balanced measure when both false positives and false negatives are important.

Macro average: The unweighted average of a metric (e.g., precision) across classes. It treats both classes equally regardless of their sample size.

Weighted average: The average of a metric weighted by the number of instances in each class. It reflects the overall performance on the entire test set.

These metrics are computed on the test set (450 samples) for each model. The results are presented in classification reports and confusion matrices.

3.6. Feature importance analysis

To address the interpretability limitation commonly associated with machine learning models, we conducted a feature importance analysis. This method measures the decrease in model accuracy when the values of a single feature are randomly shuffled across the test set, thereby breaking the relationship between that feature and the target variable (CHOICE). A large drop in accuracy indicates that the model relies heavily on that feature. Permutation importance is model-agnostic, allowing fair comparison across ANN, RF, DT, and KNN. The numerical results are provided in sections 4-5.

4. Modeling results

This section presents the results of applying the aforesaid methods on the dataset of this research.

4.1. Classification report and confusion matrix of ANN

Table 3 shows the results of classification conducted by ANN. According to this table, the neural network model developed for binary classification of travel mode choice between High-Speed Rail (HSR) and airplane (APL) was evaluated on a test set of 450 samples, consisting of 200 instances labeled as HSR and 250 instances labeled as APL. The model achieved an overall accuracy of 84.67%, indicating strong overall discriminative performance across the two classes. This means that out of 450 instances, the model correctly predicted 381 cases (171+210 in the confusion matrix). This level of accuracy reflects the network's ability to capture the complex, nonlinear relationships among socio-economic, trip-related, and attitudinal features used as predictors.

Looking at class-specific metrics, the model showed balanced yet slightly asymmetric behavior. For the HSR class, precision reached 0.8104, recall was 0.8550, and the F1-score stood at 0.8321. This means the model correctly identified 85.50% of actual HSR choosers (high recall) while 81.04% of the instances predicted as HSR were truly HSR (solid precision). These values indicate that among all instances predicted as HSR, 81.04% were actually HSR (precision), and among all true HSR instances, 85.50% were correctly identified (recall). The remaining 18.96% of HSR predictions were false positives, and 14.50% of true HSR instances were missed (false negatives). The F1-score of 0.8321 balances these two aspects. For the airplane class, precision was higher at 0.8787, recall was 0.8400, and the F1-score reached 0.8589. The higher precision indicates fewer false positives when predicting airplane choice, while the recall shows that 84.00% of true airplane choosers were correctly identified.

The aggregate performance metrics show small differences between macro-averaged and weighted-averaged values. The macro-average precision = 0.8445, recall = 0.8475, and F1-score = 0.8455. The weighted average precision = 0.8483, recall = 0.8467, and F1-score = 0.8470. The difference between macro and weighted F1-score is 0.0015, and the difference between macro F1-score and overall accuracy (0.8467) is 0.0012. These negligible differences indicate that class imbalance (200 HSR vs. 250 APL) does not substantially affect the aggregated metrics.

In summary, the neural network classifier exhibits good generalization with balanced class-wise performance, better precision for airplane predictions and higher recall for high-speed rail predictions. A direct quantitative comparison with traditional discrete choice models is not provided here but is suggested for future research.

Table 3. The classification report of ANN.

	Precision	Recall	F1-score	Support
HSR	0.8104	0.8550	0.8321	200
APL	0.8787	0.8400	0.8589	250
Accuracy			0.8467	450
Macro avg	0.8445	0.8475	0.8455	450
Weighted avg	0.8483	0.8467	0.8470	450

Also, the confusion matrix shown in Fig. 1 reveals more facts about the performance of ANN. The confusion matrix for the neural network model reveals a strong and well-balanced classification performance in distinguishing between High-Speed Rail (HSR, class 0) and airplane (APL, class 1) choices on the test set of 450 samples.

Out of the 200 actual HSR choices, the model correctly classified 171 instances (true positives), achieving a recall of 85.50%, meaning it successfully captured the vast majority of passengers who preferred high-speed rail. Only 29 HSR choosers were misclassified as airplane users (false negatives), indicating relatively low omission error for this class. On the airplane side, out of 250 true airplane preferences, the model correctly identified 210 instances (true negatives), corresponding to a recall of 84.00%. The 40 false positives (actual HSR labeled as airplane) suggest the model is somewhat conservative when predicting airplane, but still maintains very good coverage.

Overall, the matrix shows 381 correct predictions (171 + 210) out of 450, yielding the reported 84.67% accuracy. The off-diagonal elements are relatively small and nearly symmetric (29 vs. 40), which confirms the absence of severe class bias despite the slight imbalance in the test set (200 vs. 250).

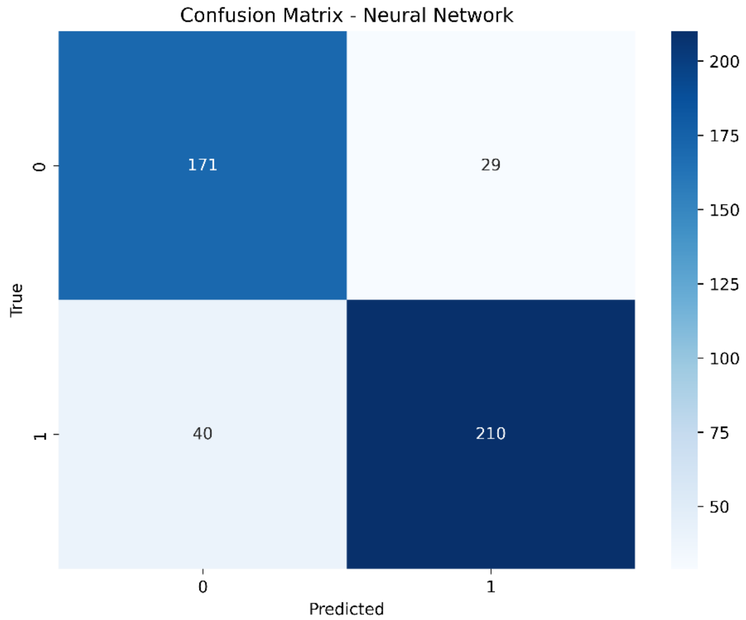


Fig. 1. The confusion matrix of ANN.

4.2. Classification report and confusion matrix of RF

The RF model attained an overall accuracy of 82.44%, reflecting predictive capability in separating the two transport modes using the input features encompassing socio-economic, trip attributes, and attitudinal variables.

Class-level performance metrics indicate reasonably balanced results with a slight edge toward the majority class (airplane), such that the precision difference between airplane (0.8490) and HSR (0.7951) is 0.0539, and the recall difference is 0.0170 (0.8320 vs. 0.8150). For the HSR class, precision was 0.7951, recall reached 0.8150, and the F1-score stood at 0.8049. This shows that the model recovered 81.50% of actual HSR preferences, while 79.51% of predictions labeled as HSR were correct; that indicates reliable performance on the minority class. For the airplane class, precision was higher at 0.8490, recall was 0.8320, and the F1-score reached 0.8404. The elevated precision demonstrates fewer false positives when predicting airplane choice, and the recall confirms that 83.20% of true airplane users were correctly identified.

The aggregate metrics reinforce the model's consistency across classes. The macro-average (unweighted) yielded precision of 0.8221, recall of 0.8235, and F1-score of 0.8227, confirming that performance is not heavily skewed toward the larger class despite the modest imbalance. The weighted average, which accounts for support, produced precision of 0.8250, recall of 0.8244, and F1-score of 0.8246, values that closely mirror the overall accuracy and attest to robust generalization. Overall, the classifier delivers discriminative power with an accuracy of 82.44% and class-wise F1-scores between 0.805 and 0.840. The precision difference (0.7951 vs. 0.8490) suggests a 5.4% higher false positive rate for HSR predictions. This pattern may reflect underlying behavioral biases, though further analysis with revealed preference data would be required to confirm. The results of RF classification are shown in Table 4.

Table 4. The classification report of RF.

	Precision	Recall	F1-score	Support
HSR	0.7951	0.8150	0.8049	200
APL	0.8490	0.8320	0.8404	250
Accuracy			0.8244	450
Macro avg	0.8221	0.8235	0.8227	450
Weighted avg	0.8250	0.8244	0.8246	450

As evident in Fig. 2, the matrix reveals that the model achieved a robust overall accuracy of approximately 82.4%, successfully predicting 371 out of 450 total samples. This performance is underpinned by 163 True Negatives for the HSR class and 208 True Positives for the airplane class. Regarding classification errors, the model exhibited a balanced error profile with 37 False Positives (Type I errors) and 42 False Negatives (Type II errors) (the slightly higher frequency of False Negatives suggests a marginal challenge in detecting the airplane class relative to HSR). The class imbalance ratio in the test set is 1.25. The false positive rate for HSR is 0.185, and for APL is 0.168, yielding a ratio of 1.10, close to the class imbalance ratio. In conclusion, the RF model demonstrates stable and reliable predictive capabilities within this context.

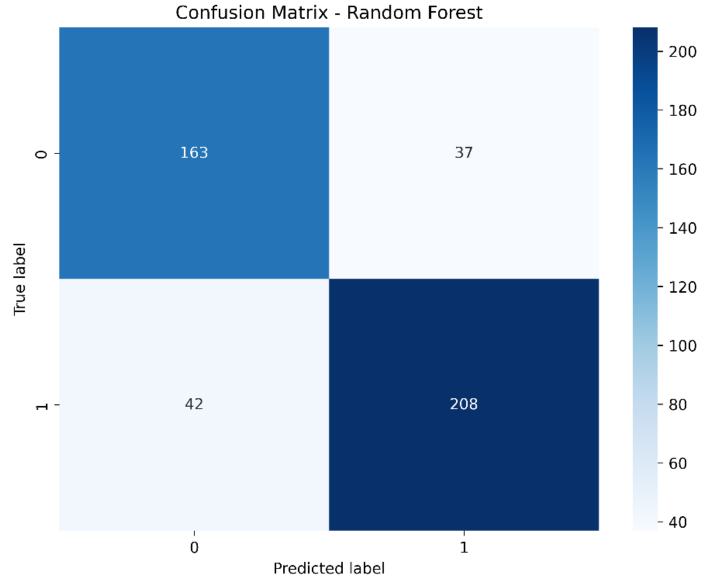


Fig. 2. The confusion matrix of RF.

4.3. Classification report and confusion matrix of DT

The DT model achieves an overall Accuracy of 75.56%. HSR class exhibits a higher Recall (0.7950) compared to its Precision (0.6974). The recall for HSR is 0.7950, while precision is 0.6974, indicating a false positive rate of 30.26% for HSR predictions. The resulting F1-score of 0.7430 represents a balance between these two metrics. Conversely, the airplane class demonstrates superior Precision (0.8153) but a relatively lower Recall (0.7240). This indicates that when the model predicts a sample as airplane, it is highly likely to be correct, although it misses approximately 27.6% of actual airplane cases. The F1-score for airplane (0.7669) is slightly higher than that of HSR. The macro-averaged F1-score (0.7550) treats both classes equally regardless of their support size. Since this value is close to the overall accuracy, it suggests that the model's performance is relatively stable and does not suffer significantly from the slight class imbalance (200 vs. 250). The weighted F1-score (0.7563) accounts for the sample distribution. The proximity between the macro and weighted averages further confirms a consistent performance across both categories. In conclusion, the model demonstrates a robust predictive capability with an F1-score exceeding 0.74 for both classes. There is a clear trade-off between the two classes. Precision for APL (0.8153) exceeds that for HSR (0.6974) by 0.1179, while recall for HSR (0.7950) exceeds that for APL (0.7240) by 0.0710. The results of DT classification are shown in Table 5.

Table 5. The classification report of DT.

	Precision	Recall	F1-score	Support
HSR	0.6974	0.7950	0.7430	200
APL	0.8153	0.7240	0.7669	250
Accuracy			0.7556	450
Macro avg	0.7563	0.7595	0.7550	450
Weighted avg	0.7629	0.7556	0.7563	450

Also, as Fig. 3 suggests, the matrix reveals 159 true negatives (correctly identified HSR instances), 181 true positives (correctly identified airplane instances), 41 false positives (HSR misclassified as airplane), and 69 false negatives (airplane misclassified as HSR), yielding an overall accuracy of 75.56%. For the HSR class (Class 0), the model achieves a recall of 79.5% and precision of 69.7%, indicating it successfully captures the majority of HSR instances but occasionally misidentifies them as airplane. Conversely, the airplane class (Class 1) demonstrates superior precision at 81.5% with a recall of 72.4%, suggesting the model is more conservative when predicting airplane(when it does, it is highly likely to be correct, though some actual airplane cases are missed). The confusion matrix reveals that Correct predictions (159 + 181 = 340) outnumber misclassifications (41 + 69 = 110) by a ratio of 3.09:1, which confirms the model's overall predictive reliability.

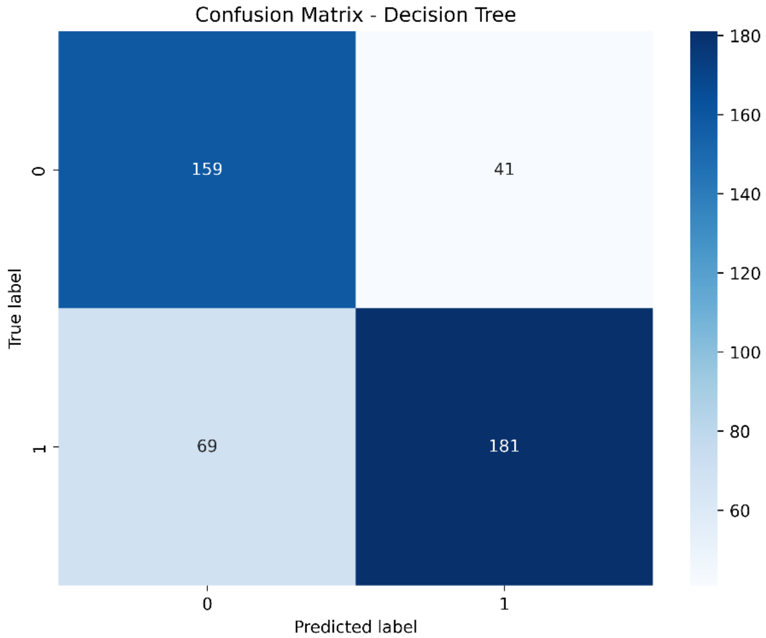


Fig. 3. The confusion matrix of DT.

4.4. Classification report and confusion matrix of KNN

The KNN model achieves an overall Accuracy of 78.44%, indicating good performance across the entire dataset. The HSR class exhibits balanced values for Precision and Recall (0.7225 and 0.7300, respectively). This convergence suggests that the model performs without significant bias in identifying HSR instances (the difference between precision and recall for HSR is 0.0075). In contrast, the airplane class demonstrates superior performance, with Precision of 0.7931, Recall of 0.8280, and F1-score of 0.8102. This indicates that the model possesses high proficiency in identifying airplane patterns. The macro-average (Macro avg), which assigns equal weight to each class, shows an F1-score of 0.7804, closely aligned with the overall Accuracy (78.44%). The macro F1-score (0.7804) and weighted F1-score (0.7837) differ by 0.0033, demonstrating that the model maintains good stability against the slight class imbalance in data distribution (200 vs. 250 samples). The weighted F1-score (Weighted avg) at 0.7837 also shows close alignment with the macro-average, further confirming the model's consistent performance across both categories. In conclusion, the model demonstrates reliable predictive capability, with F1-scores exceeding the 75% threshold for both classes. The classification report results for KNN are presented in Table 6.

Table 6. The classification report of KNN.

	Precision	Recall	F1-score	Support
HSR	0.7225	0.7300	0.7506	200
APL	0.7931	0.8280	0.8102	250
Accuracy			0.7844	450
Macro avg	0.7828	0.7790	0.7804	450
Weighted avg	0.7839	0.7844	0.7837	450

The KNN matrix in Fig 4 reveals146 true negatives (correctly identified HSR instances), 207 true positives (correctly identified airplane instances), 54 false positives (HSR misclassified as airplane), and 43 false negatives (airplane misclassified as HSR), yielding an overall accuracy of 78.44%. For the HSR class (Class 0), the model achieves a recall of 73.0% and precision of 77.2%, indicating it successfully captures the majority of HSR instances with relatively balanced performance between capturing true instances and avoiding false classifications. Conversely, the airplane class (Class 1) demonstrates strong performance with precision at 79.3% and recall at 82.8%, suggesting the model is particularly proficient at identifying airplane patterns, both catching most actual airplane cases and maintaining high confidence when it makes such predictions. The diagonal dominance in the confusion matrix, where correct predictions (146 and 207) significantly outnumber misclassifications (54 and 43), confirms the model's overall predictive reliability. The KNN model shows improved accuracy compared to the DT (78.44% vs. 75.56%), with notably lower misclassification rates for airplane (43 vs. 69 false negatives).

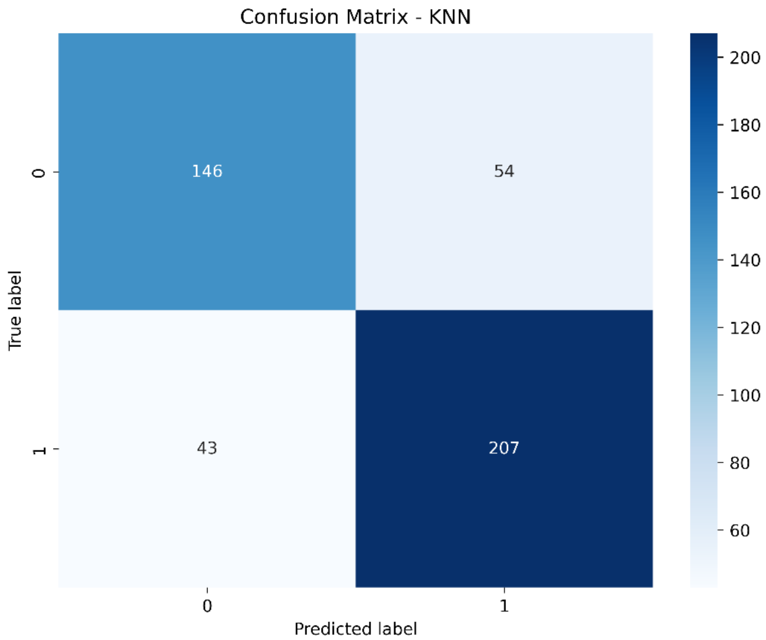


Fig. 4. The confusion matrix of KNN.

4.5. Feature importance analysis

Fig. 5 presents the permutation importance results for the top 15 features for the superior model, ANN. Based on this diagram, CST_APL (airplane ticket cost) was the most influential predictor, with a mean decrease in accuracy of 13.87% when shuffled. CST_HSR (HSR ticket cost) ranked second, with an importance of 11.71%. Travel time variables ranked third and fourth: TIME_APL (6.08%) and TIME_HSR (5.45%). The fifth most important feature was LYAPLCST (cost of air travel in the past year) at 3.76%. Household income (HHINCOME, 3.51%) and personal income (INCOME, 3.43%) showed moderate importance, followed by ATEFFRD (affordability for APL tickets, 3.18%), HSRMRWFRQ (likelihood of choosing HSR if frequency exceeds APL, 3.17%), and ENVEFFCTIMP (impact of environmental variables on mode choice, 3.16%). The remaining features, including HSRSFCWFRQ (3.16%), SFTCN (safety concerns, 3.08%), WKFRQ (weekly service frequency importance, 3.08%), ENVDDCS (environmental protection importance, 3.02%), and MRCSTLWPLT (willingness to pay more for reduced pollution, 2.90%), all showed importance values below 3.5%. Overall, the top four features (CST_APL, CST_HSR, TIME_APL, TIME_HSR) account for a total importance of 34.87%, indicating that travel cost and travel time dominate mode choice predictions in this dataset.

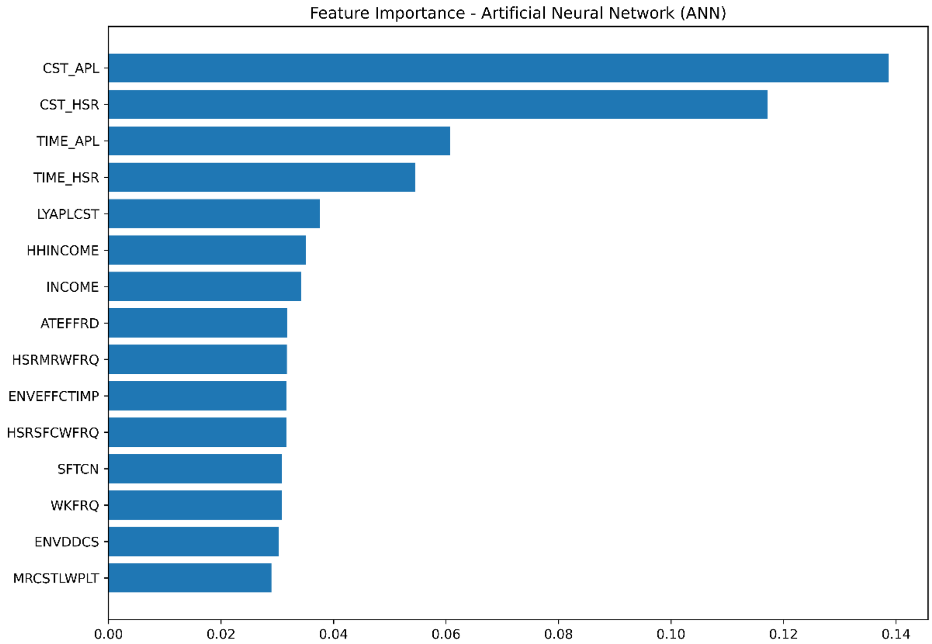


Fig. 5. The feature importance of ANN.

5. Discussion

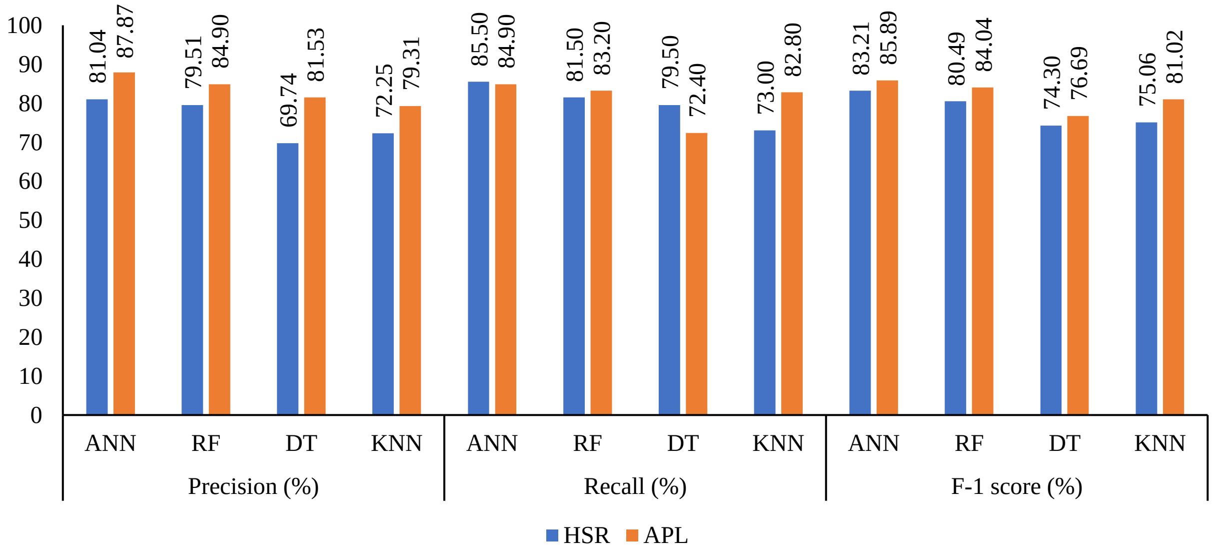
In the previous section, numerical results of different models were provided by the classification report and confusion matrix. This section is an attempt to address a major question: "Which model could outperform others?" and provide numerical insights into answering it.

The comparative analysis of machine learning models for travel mode choice classification reveals distinct performance characteristics across the four algorithms evaluated. The ANN emerged as the top performer with an overall accuracy of 84.67%, demonstrating capability in capturing complex non-linear relationships between socio-economic, trip-related, and attitudinal predictors. The RF model followed with 82.44% accuracy, showing robust predictive capability with relatively balanced class-wise performance, though slightly favoring the majority class. The KNN model achieved 78.44% accuracy with good performance on the airplane class (F1-score of 0.8102), suggesting pattern recognition for air travel preferences. Meanwhile, the DT model recorded 75.56% accuracy, exhibiting a distinct trade-off pattern characterized by higher recall for HSR (0.7950) paired with lower precision (0.6974), indicating a classification approach for rail choice at the cost of increased false positives.

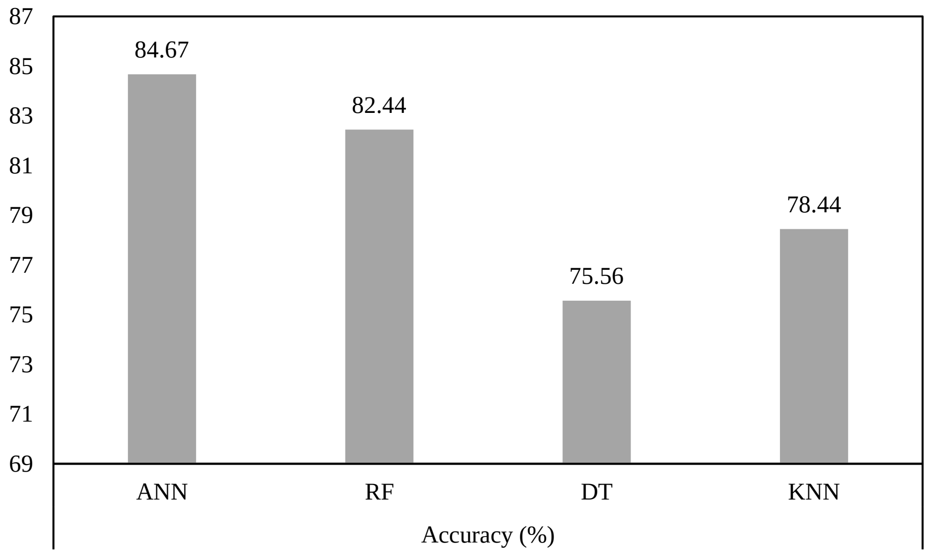
Class-specific analysis across all models consistently demonstrated higher precision for airplane predictions, reflecting fewer false positives when identifying air travel choices, while recall metrics were generally superior for HSR classifications. The proximity between macro-averaged and weighted F1-scores in all models confirmed consistent performance despite the slight class imbalance (200 HSR vs. 250 airplane samples).

Overall, the findings suggest that ensemble methods and neural network architectures offer advantages over simpler decision-based algorithms for modeling multimodal transport choices, with ANN particularly excelling in capturing the intricate interplay between attitudinal variables and mode selection behavior. These results demonstrate the predictive performance of machine learning models on this dataset. A quantitative comparison with traditional discrete choice models is not provided here but is recommended for future research.

Fig. 6a and 6b depict frameworks in which the numerical indicators of these methods are presented, hence a better numerical comparison and inference can be made.



(a)



(b)

Fig. 6. Models evaluating mode choice: (a) Precision, Recall and F-1 Score for ML, and (b) Precision, Recall and Accuracy for ML.

6. Conclusion

This study investigated the application of machine learning algorithms for mode choice modeling between high-speed rail (HSR) and airplane in Iran's transportation network, using stated preference data collected from 100 respondents across 18 travel scenarios. The comparative analysis of four machine learning models, ANN, RF, KNN, and DT, revealed that ANN achieved the highest overall accuracy (84.67%), followed by RF (82.44%), KNN (78.44%), and DT (75.56%). This ranking indicates that ensemble methods and neural network architectures are more effective in capturing complex nonlinear relationships in this dataset. Class-specific analysis showed higher precision for airplane predictions across all models, while recall metrics were generally superior for HSR classifications, suggesting that distinct psychological and attitudinal factors influence rail preferences.

- This research addresses three critical gaps in mode choice literature by being among the first Iranian studies to apply machine learning for HSR-airplane competition, successfully incorporating psychological variables (fear of flying, environmental awareness, willingness to pay) alongside socio-economic predictors, and providing algorithm evaluation using standardized metrics.
- While this study demonstrates ML's potential, limitations such as reliance on stated preference data, focus on a single corridor (Tehran-Mashhad), and limited sample size warrant future research integrating revealed preference data, multi-corridor analysis, and larger diverse datasets. Also, since the study has focused exclusively on a single corridor (Tehran-Mashhad), results may not be transferable to other routes with different distance, demographic, or competitive characteristics. Finally, psychological variables were measured using single questions rather than validated multi-item scales, which may not capture the full complexity of attitudes such as fear of flying or environmental concern.
- Regarding generalizability, the findings of this study are derived from stated preference data collected along a single corridor (Tehran-Mashhad, approximately 900 km) in Iran, a country where high-speed rail is not yet operational. While this setting provides a valuable developing-country perspective, caution should be exercised when extrapolating the results to other contexts. Several factors may limit generalizability, such as the distance and travel time values that are specific to this corridor, different income levels, price sensitivity, and modal availability across countries, psychological factors, and the absence of actual HSR service means respondents' stated preferences may differ from revealed behavior once HSR is implemented. Nevertheless, the relative ranking of model performance (ANN > RF > KNN > DT) and the dominance of cost and time as predictors suggest that certain patterns may transfer. Future research should apply the same modeling framework to multiple corridors and different developing or developed nations to test the robustness of our conclusions.
- A point that could be considered as a potential future direction is using metaheuristic algorithms, such as Water Cycle Algorithm (WCA), Wild Horse Optimization (WHO), Particle Swarm Optimization (PSO), genetic algorithm (GA), and other known algorithms for tuning hyperparameters [43-45]. Also, exploring hybrid models that embed logit-based interpretability into machine learning frameworks could be considered. In these approaches, such as utility-consistent neural networks or two-stage approaches, ML identifies nonlinear patterns and logit provides behavioral coefficients.

Statements & Declarations

Author contributions

Mohammad Feli: Conceptualization, Methodology, Validation, Formal analysis, Writing – Original Draft, Writing-Review & Editing.

Ali Naderan: Conceptualization, Methodology, Formal analysis, Project administration, Supervision, Writing - Review & Editing.

Mahmoud Saffarzadeh: Conceptualization, Methodology, Project administration, Supervision, Writing - Review & Editing.

References

[1] Banyong, C., Hantanong, N., Wisutwattanasak, P., Champahom, T., Theerathitichaipa, K., Seefong, M., Ratanavaraha, V., Jomnonkwao, S. A machine learning comparison of transportation mode changes from high-speed railway promotion in Thailand. Results in Engineering, 2024; 24: 103110. doi:10.1016/j.rineng.2024.103110.

[2] Cao, W., Chen, Z., Shi, F., Xu, J. Analysis of travel mode choice behavior between high-speed rail and air transport utilizing large-scale ticketing data. Transportation Research Record: Journal of the Transportation Research Board, 2024; 2679: 1344–1358. doi:10.1177/03611981241270169.

[3] Li, X., Wang, Y., Wu, Y., Chen, J., Zhou, J. Modeling Intercity Travel Mode Choice with Data Balance Changes: A Comparative Analysis of Bayesian Logit Model and Artificial Neural Networks. Journal of Advanced Transportation, 2021; 2021: 9219176. doi:10.1155/2021/9219176.

[4] Banyong, C., Hantanong, N., Nanthawong, S., Se, C., Wisutwattanasak, P., Champahom, T., Ratanavaraha, V., Jomnonkwao, S. Machine learning-based analysis of travel mode preferences: Neural and boosting model comparison using stated preference data from Thailand's emerging high-speed rail network. Big Data and Cognitive Computing, 2025; 9: 155. doi:10.3390/bdcc9060155.

[5] Zhao, X., Yan, X., Yu, A., Van Hentenryck, P. Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. Travel Behaviour and Society, 2020; 20: 22–35. doi:10.1016/j.tbs.2020.02.003.

[6] Ali, M. Discrete Choice Models and Artificial Intelligence Techniques for Predicting the Determinants of Transport Mode Choice—A Systematic Review. Computers, Materials and Continua, 2024; 81: 2161–2194. doi:10.32604/cmc.2024.058888.

[7] Fale, M., Wang, Y., Rupnik, B., Kramberger, T., Vizinger, T. Systematic review of transportation choice modeling. Applied Sciences, 2025; 15: 9235. doi:10.3390/app15179235.

[8] Forsythe, C. R., Arteaga, C., Helveston, J. P. The Mixed Aggregate Preference Logit Model: A Machine Learning Approach to Modeling Unobserved Heterogeneity in Discrete Choice Analysis. arXiv preprint arXiv:2402.00184, 2024; doi:10.48550/arXiv.2402.00184.

[9] Givoni, M., Dobruszkes, F. A review of ex-post evidence for mode substitution and induced demand following the introduction of high-speed rail. Transport Reviews, 2013; 33: 720–742. doi:10.1080/01441647.2013.853707.

[10] Fu, X., Oum, T. H., Yan, J. An analysis of travel demand in Japan's intercity market empirical estimation and policy simulation. Journal of Transport Economics and Policy (JTEP), 2014; 48: 97–113. doi:10.3828/jtep.2014.48.1.97.

[11] Raturi, V., Verma, A. A game-theoretic approach to analyse inter-modal competition between high-speed rail and airlines in the Indian context. Transportation Planning and Technology, 2020; 43: 20–47. doi:10.1080/03081060.2020.1701666.

[12] Xia, W., Wang, K., Zhang, A. Air transport and high-speed rail interactions in China: review on impacts of low-cost carriers, rail speed, and modal integration. Advances in Airline Economics, 2018; 7: 103–122. doi:10.1108/S2212-160920180000007007.

[13] Xia, W., Zhang, A. Air and high-speed rail transport integration on profits and welfare: Effects of air-rail connecting time. Journal of Air Transport Management, 2017; 65: 181–190. doi:10.1016/j.jairtraman.2017.06.008.

[14] Park, Y., Ha, H.-K. Analysis of the impact of high-speed railroad service on air transport demand. Transportation Research Part E: Logistics and Transportation Review, 2006; 42: 95–104. doi:10.1016/j.tre.2005.09.003.

[15] Román, C., Martín, J. Potential demand for new high speed rail services in high dense air transport corridors. International Journal of Sustainable Development and Planning, 2010; 5: 114–129. doi:10.2495/SDP-V5-N2-114-129.

[16] Raturi, V., Verma, A. Analyzing competition between High Speed Rail and Bus mode using market entry game analysis. Transportation Research Procedia, 2017; 25: 2373–2384. doi:10.1016/j.trpro.2017.05.264.

[17] Raturi, V., Verma, A. Competition between High Speed Rail and Conventional Transport Modes: Market Entry Game Analysis on Indian Corridors. Networks and Spatial Economics, 2019; 19: 763–790. doi:10.1007/s11067-018-9421-2.

[18] D'Alfonso, T., Jiang, C., Bracaglia, V. Would competition between air transport and high-speed rail benefit environment and social welfare? Transportation Research Part B: Methodological, 2015; 74: 118–137. doi:10.1016/j.trb.2015.01.007.

[19] D'Alfonso, T., Jiang, C., Bracaglia, V. Air transport and high-speed rail competition: Environmental implications and mitigation strategies. Transportation Research Part A: Policy and Practice, 2016; 92: 261–276. doi:10.1016/j.tra.2016.06.009.

[20] Wang, J., Jiao, J., Du, C., Hu, H. Competition of spatial service hinterlands between high-speed rail and air transport in China: Present and future trends. Journal of Geographical Sciences, 2015; 25: 1137–1152. doi:10.1007/s11442-015-1224-5.

[21] Edrissi, A., Javanbakht, N., Ganjipour, H. Modeling the behavior of disordered taxi drivers of Tehran for choosing passenger and destination. International Journal of Human Capital in Urban Management, 2019; 4: 69–76. doi:10.22034/IJHCUM.2019.01.08.

[22] Pagliara, F., Vassallo, J. M., Román, C. High-speed rail versus air transportation: Case study of Madrid–Barcelona, Spain. Transportation Research Record: Journal of the Transportation Research Board, 2012; 2289: 10–17. doi:10.3141/2289-02.

[23] Javanbakht, N., Mirbaha, B. Evaluating Drivers' Response to Road Hazard: A Simulation Study. Advances in Civil Engineering, 2024; 2024: 6788857. doi:10.1155/2024/6788857.

[24] Javanbakht, N., Mirbaha, B. From Traits to Speed Control: Engineering Insights from a Driving-Simulator Hazard Scenario. Contributions of Science and Technology for Engineering, 2026; 3: 10–19. doi:10.22080/cste.2025.29839.1075.

[25] Pan, J. High-Speed Rail in the US—Mode Choice Decision and Impact of COVID-19. Sustainability, 2024; 16: 4041. doi:10.3390/su16104041.

[26] Hillel, T., Bierlaire, M., Elshafie, M. Z. E. B., Jin, Y. A systematic review of machine learning classification methodologies for modelling passenger mode choice. Journal of Choice Modelling, 2021; 38: 100221. doi:10.1016/j.jocm.2020.100221.

[27] Chmielewski, J., Wójcik, M. The Use of Deep Neural Networks (DNN) in Travel Demand Modelling. Applied Sciences, 2025; 15: 11290. doi:10.3390/app152011290.

[28] Buijs, R., Koch, T., Dugundji, E. Using neural nets to predict transportation mode choice: Amsterdam network change analysis. Journal of Ambient Intelligence and Humanized Computing, 2021; 12: 121–135. doi:10.1007/s12652-020-02855-6.

[29] Tang, L., Tang, C., Fu, Q., Ma, C. Predicting travel mode choice with a robust neural network and Shapley additive explanations analysis. IET Intelligent Transport Systems, 2024; 18: 1339–1354. doi:10.1049/itr2.12514.

[30] Amoei Khorshidi, N., Zargari, S., Mirzahossein, H., Heidari, H., Waller, S. Predicting travel-time reliability in road networks: a Fitrnet-based approach–a case study of England. Scientific Journal of Silesian University of Technology. Series Transport, 2025; 126: 79–95. doi:10.20858/sjsutst.2025.126.5.

[31] Bhosle, N., Jagtap, J., Svitek, M. Machine Learning-Based Travel Mode Prediction: A Comparative Methodological Approach. In: 2025 Smart City Symposium Prague (SCSP); 2025; Prague. p. 1–5. doi:10.1109/SCSP65598.2025.11037720.

[32] Zhang, W., Luo, Z. Research on intercity travel mode recognition and network structure characteristics based on complex network and random forest classification. Scientific Reports, 2025; 15: 35339. doi:10.1038/s41598-025-19392-x.

[33] Naseri, H., Waygood, E. O. D., Patterson, Z., Alousi-Jones, M., Wang, B. Travel mode choice prediction: developing new techniques to prioritize variables and interpret black-box machine learning techniques. Transportation Planning and Technology, 2025; 48: 582–605. doi:10.1080/03081060.2024.2411611.

[34] Afandizadeh Zargari, S., Amoei Khorshidi, N., Mirzahossein, H., Heidari, H. Analyzing the effects of congestion on planning time index–Grey models vs. random forest regression. International Journal of Transportation Science and Technology, 2023; 12: 578–593. doi:10.1016/j.ijtst.2022.05.008.

[35] Kalantari, H., Sabouri, S., Brewer, S., Ewing, R., Tian, G. Machine learning in mode choice prediction as part of mpos' regional travel demand models: Is it time for change? Sustainability, 2025; 17: 3580. doi:10.3390/su17083580.

[36] Uddin, M., Anowar, S., Eluru, N. Modeling freight mode choice using machine learning classifiers: a comparative study using Commodity Flow Survey (CFS) data. Transportation Planning and Technology, 2021; 44: 543–559. doi:10.1080/03081060.2021.1927306.

[37] Ghosh, T., Nagaraj, N. Evaluating the determinants of mode choice using statistical and machine learning techniques in the Indian megacity of Bengaluru. arXiv preprint arXiv:2401.13977, 2024; doi:10.48550/arXiv.2401.13977.

[38] Dahmen, V., Weikl, S., Bogenberger, K. Interpretable machine learning for mode choice modeling on tracking-based revealed preference data. Transportation Research Record: Journal of the Transportation Research Board, 2024; 2678: 2075–2091. doi:10.1177/03611981241246973.

[39] Afandizadeh, S., Amoei Khorshidi, N., Mirzahossein, H., Shakoori, S. Predicting the fluctuation of travel time reliability as a result of congestion variations by bagging-based regressors. Civil Engineering Infrastructures Journal, 2024; 57: 85–101. doi:10.22059/ceij.2023.349853.1878.

[40] Jang, J. Travel-time prediction using K-nearest neighbor method with distance metric of correlation coefficient. The Open Transportation Journal, 2019; 13: 141–150. doi:10.2174/1874447801913010141.

[41] Lagos, F., Moreno, S., Yushimito, W. F., Brstilo, T. Urban Origin–Destination Travel Time Estimation Using K-Nearest-Neighbor-Based Methods. Mathematics, 2024; 12: 1255. doi:10.3390/math12081255.

[42] Afandizadeh Zargari, S., Amoei Khorshidi, N., Mirzahossein, H., Kalantari, N. Comparative approach for predicting travel time reliability (a case study of Virginia interstate). Innovative Infrastructure Solutions, 2021; 6: 229. doi:10.1007/s41062-021-00597-8.

[43] Khorshidi, N., Rezashoar, S., Amini, P., Zargari, S. A., Mirzahossein, H. Metaheuristic-optimized neural networks for travel time index prediction: a comparative study of wild horse and coot optimization algorithms. Transportation Engineering, 2025; 22: 100395. doi:10.1016/j.treng.2025.100395.

[44] Khorshidi, N., Zargari, S. A., Rezashoar, S., Mirzahossein, H. Optimizing Travel Time Reliability with XAI: A Virginia Interstate Network Case Using Machine Learning and Meta-Heuristics. Machine Learning with Applications, 2025; 21: 100709. doi:10.1016/j.mlwa.2025.100709.

[45] Zargari, S. A., Khorshidi, N. A., Mirzahossein, H., Shakoori, S., Jin, X. Travel time reliability prediction by genetic algorithm and machine learning models. Proceedings of the Institution of Civil Engineers - Transport, 2022; 177: 214–223. doi:10.1680/jtran.22.00065.