sustainability

MDPI

Article

A New Breakthrough in Travel Behavior Modeling Using Deep Learning: A High-Accuracy Prediction Method Based on a CNN

Xuli Wen [1] and Xin Chen [2,*]

1 School of Civil and Transportation Engineering, Southeast University Chengxian College, Nanjing 210088, China; wxltmjt@cxxy.seu.edu.cn
2 School of Management Science and Engineering, Nanjing University of Finance and Economics, Nanjing 210023, China
* Correspondence: msxinchen@nufe.edu.cn

Abstract: Accurately predicting travel mode choice is crucial for effective transportation planning and policymaking. While traditional approaches rely on discrete choice models, recent advancements in machine learning offer promising alternatives. This study proposes a novel convolutional neural network (CNN) architecture optimized using orthogonal experimental design to predict travel mode choice. Using the SwissMetro dataset, which represents a specific intercity travel context in Switzerland, we evaluate our CNN model's performance and compare it with traditional machine learning algorithms and previous studies. The key innovations of our study include: (1) an optimized CNN architecture designed to capture complex patterns in travel behavior data, and (2) the application of orthogonal experimental design to efficiently identify optimal hyperparameter settings. The results demonstrate that the proposed CNN model significantly outperforms logit models, support vector machines, random forests, gradient boosting, and even state-of-the-art techniques combining discrete choice models with neural networks. The optimized CNN achieves a remarkable 95% accuracy, surpassing the best-performing benchmarks by 14–25%. The proposed methodology offers a powerful tool for understanding travel behavior, improving travel demand forecasting, and informing transportation planning decisions. Our findings contribute to the growing body of literature on machine learning applications in transportation and pave the way for further advancements in this field.

Keywords: travel mode choice prediction; convolutional neural networks; machine learning; orthogonal experimental design; transportation planning

Academic Editors: Luca D'Acierno and Marilisa Botte

Received: 6 December 2024
Revised: 5 January 2025
Accepted: 16 January 2025
Published: 18 January 2025

Citation: Wen, X.; Chen, X. A New Breakthrough in Travel Behavior Modeling Using Deep Learning: A High-Accuracy Prediction Method Based on a CNN. *Sustainability* **2025**, *17*, 738. https://doi.org/10.3390/su17020738

1. Introduction

Predicting travel mode choices is a fundamental challenge in transportation planning and policymaking. Accurate predictions are essential for anticipating travel demand, assessing the impact of interventions, and optimizing transportation systems. Traditionally, this problem has been addressed using discrete choice models grounded in random utility maximization theory, such as multinomial logit models [1–3]. However, these models often rely on strong assumptions about decision-making processes and may struggle to capture complex non-linear relationships among variables. These limitations have spurred significant interest in machine learning approaches, which have emerged as a promising alternative, offering the ability to handle large datasets, uncover intricate patterns, and model complex relationships without imposing restrictive assumptions [4–6]. Extensive research has demonstrated the superiority of machine learning approaches over traditional logit models in travel mode choice prediction [7–12]. These studies have consistently demonstrated the superiority of machine learning methods over traditional logit models. For instance, Xie et al. [13] found that decision tree (DT) and neural network methods offered comparably better performances than multinomial logit (MNL) in modeling travel mode choice. Omrani [14] demonstrated that artificial neural networks (ANNs) outperformed MNL and support vector machines (SVMs) in predicting individual travel modes. Similarly, Hagenauer and Helbich [15] reported that machine learning models, including DT, random forests (RFs), ANNs, and SVMs, outperformed the MNL in terms of model accuracy. Moreover, Zhao et al. [16] found that the RM has significantly higher accuracy in capturing complex non-linear relationships than MNL and mixed logit models. Moreover, the integration of advanced deep learning techniques for travel mode prediction has become an important research frontier. Dabiri and Heaslip [17] utilized convolutional neural networks (CNNs) to analyze travel patterns from raw GPS data, showing superiority over traditional algorithms. Ma and Zhang [18] employed deep neural networks (DNN) with entity embeddings to predict travel choices, outperforming simpler neural network architectures and tree-based models. Yazdizadeh et al. [19] enhanced prediction accuracy by combining a CNN with ensemble learning strategies. Yu [20] proposed a deep learning model based on long short-term memory (LSTM) networks to extract and predict travel mode choices from time-frequency domain features of trajectories. Recent research has made significant advances in developing interpretable machine learning approaches for travel mode choice modeling. By applying rough sets theory, Cheng et al. [21] developed a framework that establishes transparent relationships between sociodemographic characteristics and transportation preferences, making the decision-making process more comprehensible. Naseri et al. [22] introduced an innovative methodology combining Gray Relational Analysis with feature importance rankings to decode complex model behaviors and identify the key determinants of travel choices. Tamim Kashifi et al. [23] integrated variable importance analysis with the SHAPs (SHapley Additive exPlanations) methodology to provide detailed insights into how specific factors influence passenger preferences and mode selection decisions. This comprehensive approach offers a nuanced understanding of the feature selection process and its implications for travel behavior prediction. Zhang et al. [24] complemented these efforts by developing an information gain-based methodology that generates readily interpretable classification rules, effectively addressing the transparency challenges commonly associated with machine learning algorithms in transportation research.

However, a critical challenge in applying deep learning to travel mode choice prediction remains: the lack of systematic approaches for architecture optimization. Current CNN architecture design primarily relies on researcher experience and trial-and-error methods, leading to suboptimal performances and inefficient development processes [25]. This study addresses these gaps by proposing a novel CNN architecture optimization framework based on orthogonal experimental design. Our approach systematically explores the architectural space while dramatically reducing the computational requirements compared to traditional optimization methods. We validate our methodology using the SwissMetro dataset, demonstrating its effectiveness in predicting travel mode choices. Therefore, our research makes four primary contributions to transportation mode choice modeling. First, we introduce a methodological framework that reduces the required experimental configurations by 99.4% while maintaining optimization quality. Second, our optimized CNN achieves a 95% prediction accuracy, representing a 14–25% improvement over other traditional supervised machine learning algorithms that have widely been used in the literature. Third, we provide quantitative insights into the relationship between architectural choices and model performance through rigorous analysis. Fourth, we establish a practical framework for implementing the CNN model in transportation planning contexts, including data preprocessing requirements, model training configurations, and evaluation metrics.

The remainder of this paper is organized as follows. Section 2 describes our methodology, including data preprocessing, CNN architecture design, and the orthogonal experimental design approach. Section 3 presents our results, comparing our model's performance with traditional methods and previous studies. Section 4 discusses the implications of our findings and concludes with recommendations for future research.

2. Materials and Methods

2.1. Material Collection

This study utilizes a dataset derived from a travel mode choice survey conducted to estimate the hypothetical demand for SwissMetro, a proposed innovation to the Swiss transport system aimed at connecting major urban centers in Switzerland [26]. The SwissMetro Stated Preference (SP) survey collected data from 1192 respondents (441 rail travelers and 751 car users), each with nine choices. Respondents were asked to choose a travel mode from a set of intercity travel alternatives, including trains (TRAIN), the SwissMetro (SM), and cars (CAR), given the attributes of each mode (e.g., travel time, headway, and cost). This dataset has been used by other researchers to test the efficacy of their proposed models [27–31].

The feature variables in the dataset cover the socioeconomic attributes of travelers (e.g., age, gender), travel characteristics (e.g., trip purpose, number of luggage), and the service levels of different transportation modes (e.g., cost, time, availability). The transportation mode choice (CHOICE) is the target variable, and a CNN model will be built to predict the most likely transportation mode chosen by travelers under given conditions. The dataset contains categorical variables (trip purpose, payment method, amount of luggage, age, gender, availability) and continuous variables (cost, time), providing rich information for model training and generalization. However, this diversity and heterogeneity also necessitate careful data preprocessing and feature engineering. For simplicity, 20 key attributes are kept for further model building and analysis (see Table 1). These attributes cover multiple dimensions, including the socioeconomic attributes of travelers, travel characteristics, and transportation service levels, comprehensively characterizing travelers' transportation mode choice behavior. This can not only reduce data dimensionality and computational overhead but also improve the model's interpretability and generalization. A detailed description of the variables used in this study is presented in Table 1. In the subsequent data preprocessing section, transformations such as encoding and normalization will be performed on these attributes to adapt to the input requirements of the CNN model.

Table 1. Description of the variables used in this study.

No.	Variable	Description
1	PURPOSE	Travel purpose. 1: Commuter, 2: Shopping, 3: Business, 4: Leisure, 5: Return from work, 6: Return from shopping, 7: Return from business, 8: Return from leisure, 9: Other.
2	WHO	Who pays (0: unknown, 1: self, 2: employer, 3: half–half).
3	LUGGAGE	0: none, 1: one piece, 3: several pieces.
4	AGE	AGE captures the age class of individuals. The age-class coding scheme is of the type 1: age $\leq$ 24, 2: 24 < age $\leq$ 39, 3: 39 < age $\leq$ 54, 4: 54 < age $\leq$ 65, 5: 65 < age, 6: not known.
5	MALE	Traveler's Gender: 0: female; 1: male.

Table 1. *Cont.*

No.	Variable	Description
6	INCOME	Traveler's income per year [thousand CHF], 0 or 1: under 50; 2: between 50 and 100; 3: over 100; 4: unknown.
7	GA	Variable capturing the effect of the Swiss annual season ticket for the rail system and most local public transport. It is 1 if the individual owns a GA, zero otherwise.
8	TRAIN_AV	Train availability dummy.
9	CAR_AV	Car availability dummy.
10	SM_AV	SM availability dummy.
11	TRAIN_TT	Train travel time [minutes]. Travel times are door-to-door, making assumptions about car-based distances (1.25 × crow-flight distance)
12	TRAIN_CO	Train cost [CHF]. If the traveler has a GA, this cost equals the cost of the annual ticket.
13	TRAIN_HE	Train headway [minutes]. Example: If there are two trains per hour, the value of TRAIN HE is 30.
14	SM_TT	SM travel time [minutes] considering the future SM speed of 500 km/h
15	SM_CO	SM cost [CHF] calculated at the current relevant rail fare, without considering GA, multiplied by a fixed factor (1.2) to reflect the higher speed.
16	SM_HE	SM headway [minutes]. Example: If there are two SM per hour, the value of SM_HE is 30.
17	SM_SEATS	Seat configuration in the SM (dummy). Airline seats (1) or not (0).
18	CAR_TT	Car travel time [minutes].
19	CAR_CO	Car cost [CHF] considering a fixed average cost per kilometer (1.20 CHF/km).
20	CHOICE	Choice indicator. 0: unknown, 1: Train, 2: SM, 3: Car.

2.2. Data Preprocessing

The original data comprised 10,728 observations. After removing observations with unknown travel choices (i.e., CHOICE = 0), 10,719 records were retained. To further refine the data, we used the following preprocessing procedures:

(1) One-hot encoding: In this study, we are faced with a multi-class classification problem, where the target variable "CHOICE" consists of three mutually exclusive categories, namely, TRAIN, SM, and CAR. To ensure our models can efficiently process this categorical data, one-hot encoding of the target variable is applied. This technique generates binary indicators for each category, transforming the "CHOICE" variable into three distinct columns: TRAIN_CHOICE, SM_CHOICE, and CAR_CHOICE. For example, if a sample indicates TRAIN, the TRAIN_CHOICE column is marked as 1, with the others set to 0. This encoding technique is essential to ensure each travel mode is represented numerically and equally in the dataset. One-hot encoding preserves the fundamental nature of discrete choice modeling by avoiding artificial ordinal relationships between travel modes. Alternative methods like label encoding could introduce unintended hierarchical relationships between modes, while target encoding risks data leakage during model training.

(2) Feature Scaling: Z-score standardization was employed to normalize continuous features such as train fare and traveler's annual income. This method adjusts the data to achieve a mean of 0 and a variance of 1, standardizing the scale of these variables. This process involves calculating the mean and standard deviation for each attribute and then transforming each of them to the normalized value $Z_i = (x_i - \mu)/\sigma$, where x_i is the original value of attribute i, μ is the mean, σ is the standard deviation of attribute i, and is the normalized value. This transformation can help improve the convergence speed and stability of the CNN model [32].

(3) Data Reconstruction: To align with the input requirements of the CNN model, the preprocessed attribute data were reshaped into a two-dimensional matrix, with rows corresponding to different travel modes and columns corresponding to different attributes.

By stacking these matrices for all samples, a three-dimensional tensor was created, with dimensions corresponding to the number of samples, travel modes, and features, respectively. This approach allows CNNs to extract spatial patterns of travelers' choice behavior.

(4) Addressing Class imbalance: The targeted upsampling method was used to solve the problem of category imbalance in the dataset. First, a statistical analysis confirmed the existence of both majority and minority groups. Based on this, we increased the sample size of the minority classes represented by labels 0 and 2 using a random oversampling technique. To achieve a balance in the number of classes, we specifically selected samples from a limited number of categories by random oversampling and then replicated them. The ability to identify several categories has been improved, and this process is expected to improve overall classification performance, ensuring that attributes for all categories can be fully processed during model training.

After data preprocessing, the dataset increased from 10,719 to 27,322 samples. The dataset was randomly split into training (80%, 21,857 samples) and testing (20%, 5465 samples) sets. The training set is used for model training and validation, while the testing set is used for evaluating the model's generalization performance.

2.3. CNN Architecture Based on Orthogonal Experimental Design

The performance of a CNN can be impacted by the quality of the training data, the learning algorithms, and the model's architecture. However, the current CNN architecture design mainly depends on researchers' experience and trial-and-error, lacking systematic guiding principles. To explore the CNN architecture space more effectively, an optimization method based on orthogonal experimental design is proposed in this paper. This method determines the key structural factors affecting CNN performance and their optimal level combinations with fewer experiments while providing insights into the interactions between factors, enhancing our understanding of the intrinsic mechanisms of CNN architectures. The orthogonal experimental design offers significant advantages over traditional hyperparameter optimization methods like grid search and random search. It reduced our experimental configurations from 2048 to just 12 combinations, a 99.4% reduction in computational cost while maintaining optimization quality. This systematic approach enables a quantitative assessment of each hyperparameter's impact through range analysis, provides reproducible results, and makes CNN architecture optimization feasible with limited computational resources. The 95% accuracy achieved validates the effectiveness of this efficient methodology.

When determining the optimal CNN architecture, a set of layers was chosen to form the backbone of the model, including the input layer, convolutional layers, pooling layers, dense layers, dropout layers, and softmax layers. With the parameters of the initial convolutional layer and the three pooling layers predefined, 11 CNN factors were selected, namely, the number of convolutional layers, the number of convolutional kernels in each layer, and the dropout rate. Each factor has two levels, Level_1 indexed with "1", representing that the factor is included in the model, and Level_2 indexed with "0", representing the factor is not included in the model. An L12(211) orthogonal table was adopted, requiring only 12 experiments to preliminarily determine the optimal levels of each factor. Table 2 lists the level distribution for the factors considered.

According to the orthogonal table, 12 CNN models with different structures are designed, trained, and evaluated on the dataset. The results from Table 3 demonstrate that different CNN structural configurations have a significant impact on model performance. Model VIII achieved the highest validation accuracy of 0.91 (in bold) with a structure containing three convolutional layers, three max-pooling layers, and two dense layers. In contrast, Models VI and XI had the lowest accuracies of 0.57 and 0.55, respectively, containing only one convolutional layer, one max-pooling layer, and no dense layers. This indicates that appropriately increasing the number and complexity of convolutional and dense layers can improve CNN model performance.

Table 2. CNN structural factors and levels for orthogonal experimental design.

Factors	Level_1	Level_2
Convolutional_32	*1*	*1*
Max-pooling	*1*	*1*
Dropout (0.5)	0	1
Convolutional_64	0	1
Max-pooling	*1*	*1*
Dropout (0.5)	0	1
Convolutional_128	0	1
Max-pooling	*1*	*1*
Dropout (0.5)	0	1
Dense_64	0	1
Dropout (0.5)	0	1
Dense_32	0	1
Dropout (0.5)	0	1
Dense_16	0	1
Dropout (0.5)	0	1

Note: The factors in italics are considered the essential parts of the CNN model in this study.

Table 3. Factors and levels of the orthogonal design of L12(2^{11}) for CNN models.

Layer	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Input layer	1	1	1	1	1	1	1	1	1	1	1	1
Con2d-32	1	1	1	1	1	1	1	1	1	1	1	1
Maxpool	1	1	1	1	1	1	1	1	1	1	1	1
Dropout (F1)	1	1	1	1	1	1	0	0	0	0	0	0
Con2d-64 (F2)	1	1	1	0	0	0	1	1	1	0	0	0
Maxpool	1	1	1	1	1	1	1	1	1	1	1	1
Dropout (F3)	1	1	0	1	0	0	0	0	1	0	1	1
Con2d-128 (F4)	1	1	0	0	1	0	0	1	0	1	0	1
Maxpool	1	1	1	1	1	1	1	1	1	1	1	1
Dropout (F5)	1	1	0	0	0	1	1	0	0	1	1	0
Dense-64 (F6)	1	0	1	1	0	0	1	0	0	1	0	1
Dropout (F7)	1	0	1	0	1	0	0	0	1	1	1	0
Dense-32 (F8)	1	0	1	0	0	1	0	1	0	0	1	1
Dropout (F9)	1	0	0	1	1	0	1	1	0	0	1	0
Dense-16 (F10)	1	0	0	1	0	1	0	1	1	1	0	0
Dropout (F11)	1	0	0	0	1	1	1	0	1	0	0	1
Accuracy	0.81	0.67	0.75	0.63	0.67	0.57	0.79	**0.91**	0.76	0.78	0.55	0.69

To further analyze the impact of each factor on CNN model performance, the range analysis is conducted. The average accuracy results for each factor at different levels ($\overline{K}_1$ and $\overline{K}_0$) are shown in Table 4. The larger the R-value, the greater the impact of the level change in this factor on accuracy. Then, we can see that factors such as F2 (Conv2D-64), F4 (Conv2D-128), F6 (Dense-64), F7 (Dropout following Dense-64), F9 (Dropout following Dense-32), and

F10 (Dense-16) are identified as having a positive impact on the model performance, with the optimal level being "1", indicating they should be included in the model. Other factors including F1 (Dropout following Conv2D-32) and F11 (Dropout following Dense-16) are found to have an optimal level at "0", implying they should not be included in the model. These results suggest that incorporating additional convolutional layers (Conv2D-64 and Conv2D-128) and a dense layer (Dense-16) into the model architecture can significantly enhance model performance while removing dropout layers after Conv2D-32 and Dense-16 can further optimize the model performance.

Table 4. Range analysis (R) on the factors based on orthogonal experiments.

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11
$\overline{K}_1$	0.683	**0.782**	0.685	**0.755**	0.695	**0.742**	**0.720**	0.713	**0.727**	**0.743**	0.714
$\overline{K}_0$	**0.747**	0.648	**0.745**	0.698	**0.735**	0.688	0.710	**0.717**	0.703	0.687	**0.715**
R	0.063	0.133	0.060	0.057	0.040	0.053	0.010	0.003	0.023	0.057	0.000
Best level	0	1	0	1	0	1	1	0	1	1	0

Note: the bold numbers denote the largest ones.

Therefore, we obtained an optimized CNN architecture, shown in Figure 1. This optimized model starts with a convolutional layer equipped with 3 × 1 × 12 kernels, followed by a 2 × 2 max-pooling layer, and then the second convolutional layer with 64 × 3 × 3 kernels, followed by a third convolutional layer, containing 128 × 3 × 3 kernels, and two dense layers with 32 and 16 units. The ReLU activation function was utilized for the convolutional and dense layers, and the dropout layer was initialized with a dropout rate of 0.5 to regularize the network and mitigate overfitting.

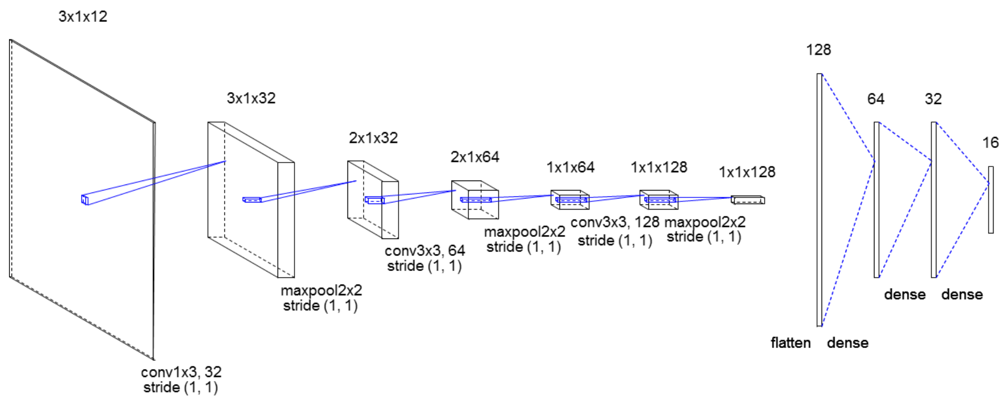


Figure 1. The optimized architecture of the CNN model.

To verify the effectiveness of the optimized CNN model, we further trained it on the complete training set and evaluated it on the test set, achieving an accuracy of 95%, confirming the superior performance of this model architecture. This demonstrates that an orthogonal experimental design can be helpful in the optimization of CNN architectures, providing a new approach to designing CNN models.

2.4. Model Training and Evaluation

After determining the optimal architecture and hyperparameters of the CNN model, the model was trained using the training set, and its performance was evaluated on the test set.

2.4.1. Training Configuration

The Stochastic Gradient Descent (SGD) optimizer and the categorical cross-entropy loss function were used to train the CNN model. The SGD optimizer updates the model parameters by computing the gradients for each mini batch of data, and it is a widely used optimization algorithm. The categorical cross-entropy loss function measures the difference between the predicted class distribution and the true distribution, making it suitable for multi-class classification problems. The batch size was set to 256, meaning that the model parameters are updated using 256 samples at each iteration. The initial learning rate was set to 0.1, controlling the step size of parameter updates. The model was trained for 500 epochs on the training set, which means the model passes through the entire training set 500 times. Moreover, to prevent overfitting, we adopted the following measures:

(a) Regularization: an L2 regularization method is included in the loss function to constrain the magnitude of model parameters, ensuring smaller parameter values.
(b) Batch Normalization: Batch Normalization layers are added after the convolutional and dense layers to normalize each mini batch of data, accelerating the training process and improving model stability.

2.4.2. Evaluation Metrics

To comprehensively evaluate the performance of the optimized CNN model, the following evaluation metrics were calculated on the test set:

(a) Accuracy: The proportion of correctly predicted samples out of the total number of samples.
(b) Precision: For each class, the proportion of samples predicted as that class and actually belonging to that class out of the samples predicted as that class.
(c) Recall: For each class, the proportion of samples predicted as that class and actually belonging to that class out of the samples actually belonging to that class.
(d) F1-score: The harmonic mean of precision and recall, considering both the accuracy and completeness of the classification.

These metrics assess the classification performance of the model from different perspectives. Accuracy reflects the overall classification correctness, while precision and recall focus on the prediction quality for each class. The F1 score balances precision and recall.

2.4.3. Training Process Monitoring

During the model training process, the loss function values and accuracy on the training set and validation set are monitored in real time to track the learning progress and generalization ability of the model. By visualizing the training curves, overfitting or underfitting can be detected and hyperparameters or appropriate measures can be adjusted accordingly. Moreover, the Early Stopping technique was employed to avoid overfitting. When the performance on the validation set no longer improved for a certain number of epochs, the training was stopped and the model parameters with the best performance were saved. This approach prevents the model from excessively fitting the training data and enhances its generalization ability on unseen data.

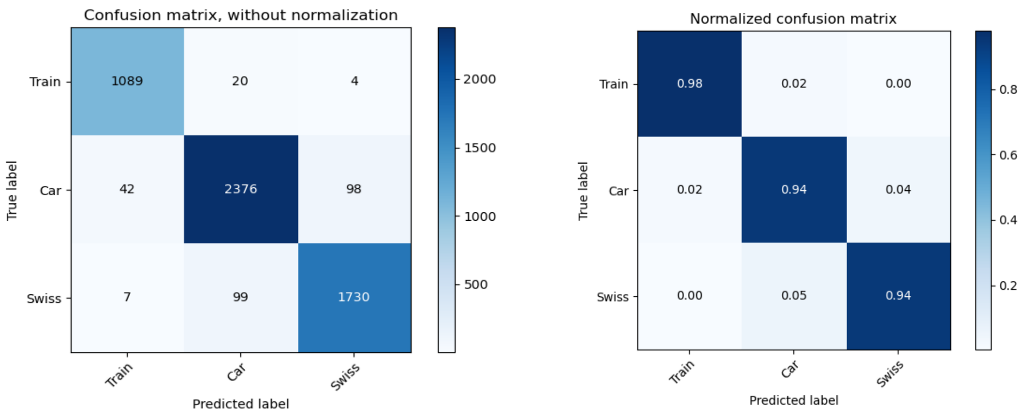
3. Results and Discussion

3.1. CNN Model Performance

The performance metrics for the CNN model are given in Table 5. From the Table, we can see that the optimized CNN model demonstrates a high-quality performance for all three classes. For TRAIN, the model achieves a precision of 0.96, recall of 0.98, and F1-score of 0.97. Concerning CAR and SM, it also exhibits impressive performance, with precision and recall values close to 0.95 and F1-scores of 0.95, suggesting that the model performs consistently across all categories. The confusion matrix in Figure 2 also provides detailed prediction results from the optimized CNN model for each travel mode. The unnormalized confusion matrix, as shown in Figure 2a, shows that on the test set, the model correctly predicted 1089 samples of Train, 2376 samples of CAR, and 1730 samples of SM. The normalized confusion matrix, as shown in Figure 2b, represents the prediction probabilities for each mode. The model achieved an overall accuracy of 95%, indicating its excellent performance overall. However, the confusion matrices also revealed some instances of misclassification, such as 42 samples of CAR being incorrectly predicted as TRAIN and 99 samples of SM being incorrectly predicted as CAR. Future improvements should be made to further optimize the model's representation. The receiver operating characteristic curve is presented in Figure 3; the total area under the curve (AUC) further demonstrates the higher model performance.

Table 5. Performance evaluation results of the CNN model.

Class	Precision	Recall	F1-Score	Accuracy
TRAIN	0.96	0.98	0.97	
CAR	0.95	0.94	0.95	0.95
SM	0.94	0.94	0.94	



(**a**) confusion matrix without normalization (**b**) normalized confusion matrix

Figure 2. Confusion matrix results of the optimized CNN.

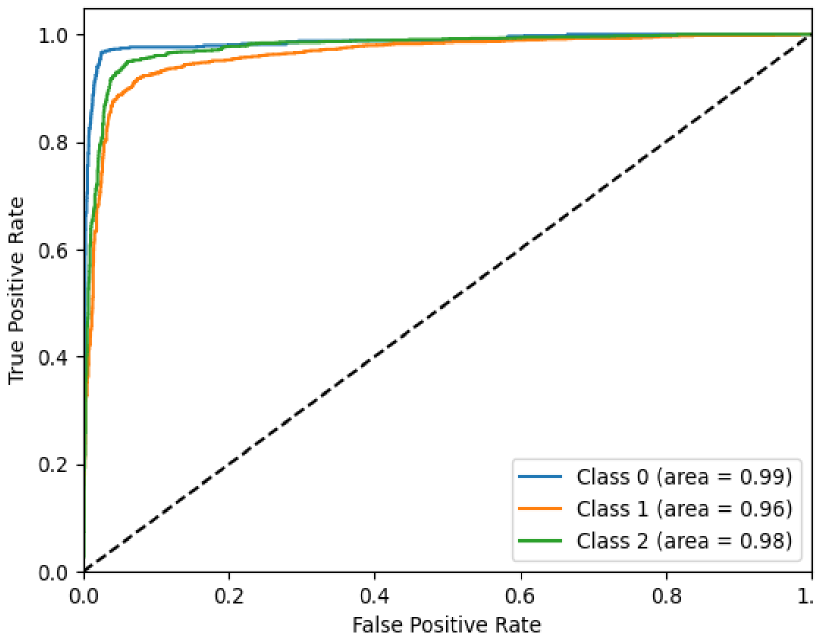


Figure 3. The AUC-ROC curve of the optimized CNN.

Figures 4 and 5 compare the test and training scores for varying epochs. The optimized model triggered the early stopping mechanism at epoch 315, preventing overfitting and indicating that the model achieved its best performance during the training process. As is shown, the training accuracy goes up by increasing the number of epochs and peaks at almost 98%. The test accuracy also goes up along with the increasing trend in the training accuracy. This in turn verifies that the overfitting problem does not occur in our optimized CNN model. The training and validation loss curves also demonstrate good model generalization without significant overfitting.

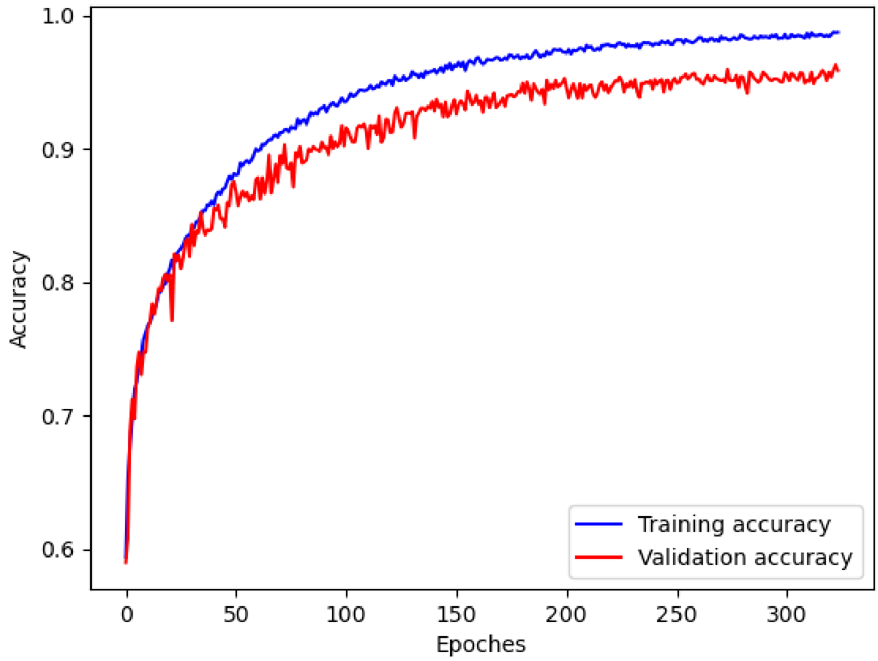


Figure 4. Comparing the training and test accuracy of the optimized CNN.

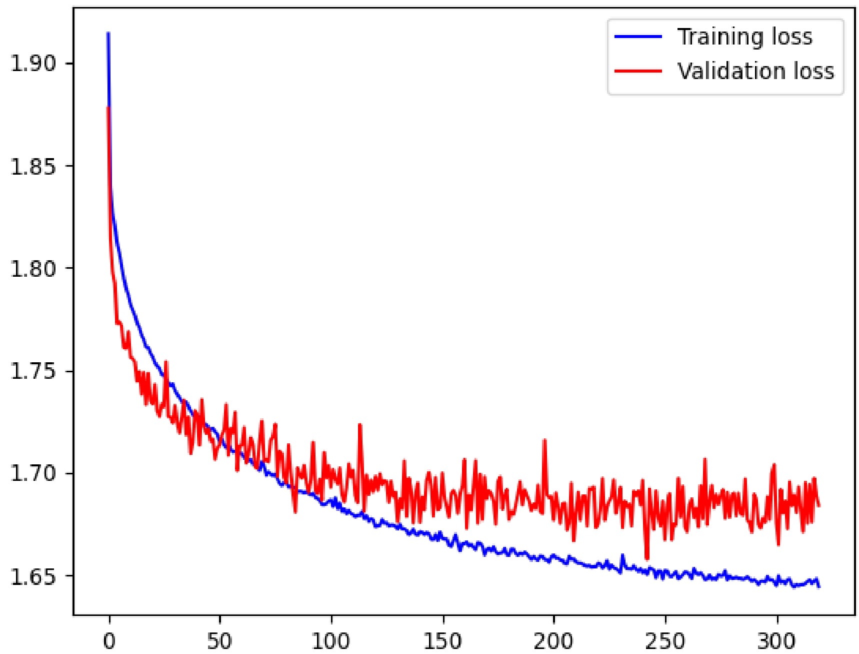


Figure 5. Comparing the training and test loss of the optimized CNN.

Furthermore, we conduct a sensitivity analysis to understand how input feature changes affect predictions. The result shown in Figure 6 reveals the relative importance of different features. First, travel time ('tt') and cost showed the highest sensitivity, with changes in these features having the largest impact on predictions. Second, service characteristics (headway and season ticket availability) demonstrated moderate influence. Third, socioeconomic variables (income and age) showed lower but consistent effects. The results provide valuable insights for transportation planners, highlighting which factors most strongly influence travel mode choice decisions.

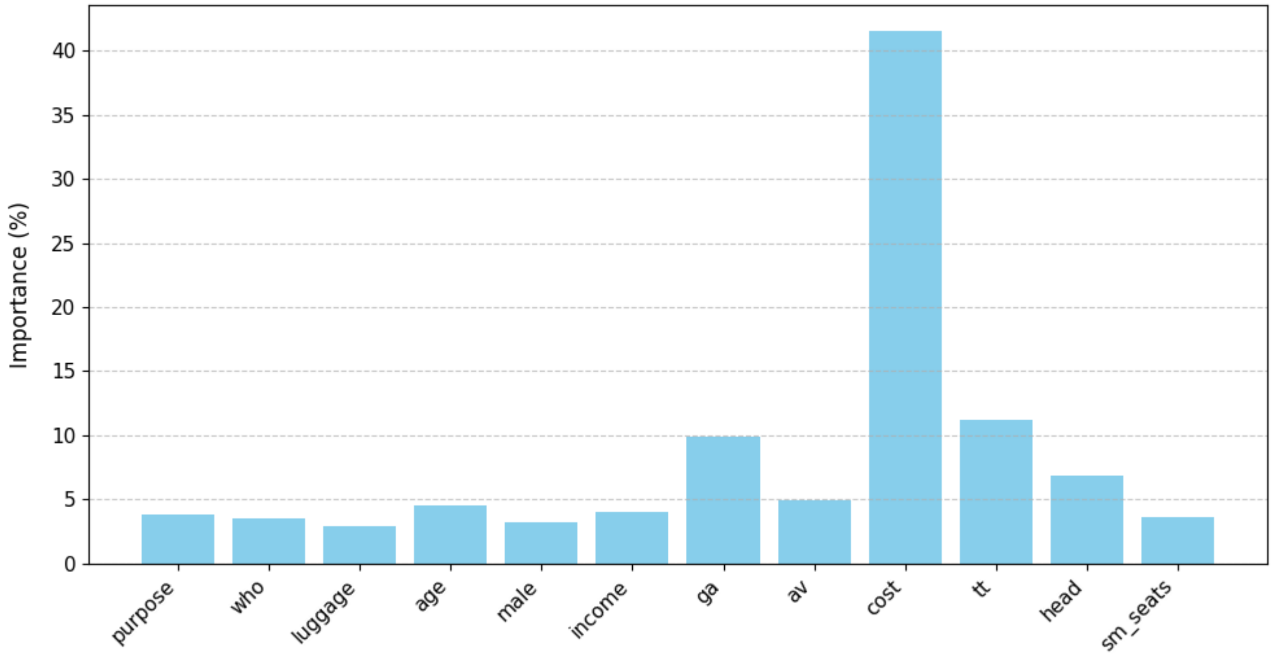


Figure 6. Feature importance ranking by sensitivity analysis.

Overall, the CNN model exhibits high accuracy and stability in the travel mode choice prediction. Both the evaluation metrics and training process prove that the CNN model can effectively learn and predict travelers' behavior, providing strong support for its advantageous application in transport demand analysis and management.

3.2. Comparison with Other Machine Learning Models

A comparative analysis of convolutional neural networks (CNNs) and five classical machine learning models was conducted to evaluate their performance in handling the task of predicting travel mode choice. These algorithms include Multiple Logistic Regression (MNL), support vector machines (SVM), random forest (RF), Gradient Boosting Decision Trees (GBDT), and Extreme Gradient Boosting (XGBoost). To ensure the fairness and consistency of the comparison, all models are trained and tested on the same dataset as used in the CNN model. Through grid search and 5-fold cross-validation, the optimal hyperparameter combinations for each model were identified, as shown in Table 6.

Based on Table 6, it can be seen that the optimized CNN model outperformed all the other models in terms of every type of evaluation metric, which further highlights its superiority in predicting travel mode choice. Specifically, the average accuracy of the optimized CNN model is 95%, while the average accuracy of the GBDT model is only 78% and MNL has the lowest accuracy of 66%. This significant difference underscores the power of CNNs in capturing complex patterns and features of the input data, implying its effectiveness in understanding and predicting travel mode, offering valuable tools for transport planning and policymaking.

Table 6. Comparison of CNN's performance with the classical machine learning algorithms.

Model	Hyperparameter	Optimal HP Values	Test Accuracy	Average Precision	Average Recall	Average F-Score
MNL	estimation method, maxiter	[bfgs, 100]	0.66	0.66	0.52	0.55
SVM	C, gamma, kernel	[10, 0.1, rbf′]	0.73	0.70	0.65	0.67
RF	max_depth of tree, min_samples_split, n_estimators	[None, 2, 100]	0.77	0.76	0.68	0.71
GBDT	max_depth of tree, max_features, subsample, learning_rate	[9, 0.6, 0.8, 0.1]	0.78	0.77	0.70	0.72
XGBoost	max_depth of tree, learning_rate, n_estimators	[7, 0.1, 200]	0.78	0.76	0.70	0.73
The optimized CNN	Already discussed above	NA	**0.95**	**0.95**	**0.96**	**0.95**

Note: HP denotes hyperparameter, and NA denotes not applicable [17]. The bold numbers denote the largest ones.

3.3. Comparison with Previous Studies

To demonstrate the superiority of our optimized CNN model in predicting travel mode choice, we compare our results with previous studies that have used the same SwissMetro dataset. The study of Sifringer et al. [31] developed a new method to improve the accuracy of choice behavior prediction in the transportation field by combining traditional discrete choice models and neural networks (NNs) techniques. Han et al. [28] proposed the TasteNet-MNL model to enhance the representation of heterogeneity and improve the interpretability of discrete choice models. Additionally, we compare our results with more recent studies that have employed deep learning techniques (DNN) [30]. Table 7 presents the prediction performance of our best CNN model alongside the highest accuracy reported in these selected studies.

Table 7. Performance comparison with other machine learning models in previous studies.

Model	Test Accuracy (%)
NN [31]	81
DNN(RRRS) [30]	72
TasteNet-MNL [28]	70
The optimized CNN model in this study:	95

As can be seen from Table 7, our optimized CNN architecture significantly outperforms the models in the previous studies, achieving an improvement in accuracy of approximately 14%, 23%, and 25% compared to the work of Sifringer et al. [31], Han et al. [28], and Nam and Cho [30], respectively. This remarkable accuracy can be attributed to the effective configuration of our optimized CNN architecture based on the orthogonal experimental design, which ensures the properties of CNNs are able to learn and extract intricate patterns and dependencies from the training data. A comparison with other machine learning models further validates the powerful prediction capabilities of the optimized CNN model. However, it should be noted that while the proposed CNN model demonstrates a superior performance on the SwissMetro dataset, representing a specific intercity travel context in Switzerland, it cannot automatically be generalized to significantly different transportation contexts without further validation.

4. Conclusions

This study investigates the application of convolutional neural networks (CNNs) in predicting travel mode choice. Through the orthogonal experimental design method, we obtained an optimized CNN architecture and configuration, which was then validated on the well-known SwissMetro dataset and its performance was compared with traditional machine learning algorithms from previous studies. The results indicate that our optimized CNN model can significantly outperform other classic machine learning algorithms such as MNL, SVM, RF, GBDT, and XGBoost. Moreover, our optimized CNN model surpassed previous studies using the same dataset, including studies that explored combinations of discrete choice models with neural networks and deep learning techniques. These findings highlight the outstanding strength of CNNs in capturing complex patterns and attributes, providing new perspectives and solutions for travel behavior modeling and transportation planning. Specifically, our proposed model can help evaluate the potential modal shifts from introducing new transportation options, similar to the SwissMetro case. In demand forecasting, the model's ability to process complex travel behavior patterns enables more accurate predictions of ridership and capacity requirements. Transportation planners can use these insights to optimize service frequency, adjust pricing strategies, and make evidence-based infrastructure investment decisions.

Despite the outstanding performance of the optimized CNN model, the scale of the specific SwissMetro dataset may limit the full potential of the CNN framework in different transportation contexts, particularly in environments with different infrastructure qualities, cultural contexts, and choice dynamics. Future research should focus on exploring larger and more diverse datasets from different geographic regions to enhance its scalability and developing innovative CNN architectures that include attention mechanisms for better feature interaction modeling. Moreover, developing more interpretable CNN architectures would be helpful to improve model transparency and provide valuable insights for real-world transportation planning and policymaking processes. These extensions would help establish the broader applicability of our CNN approach while maintaining its current advantages in predictive accuracy.

Author Contributions: X.W.: investigation, methodology, writing—original draft. X.C.: conceptualization, supervision and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in this study.

Data Availability Statement: Data can be provided as needed.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Train, K.E. *Discrete Choice Methods with Simulation*; Cambridge University Press: Cambridge, UK, 2009.
2. Hensher, D.A.; Rose, J.M.; Greene, W.H. *Applied Choice Analysis*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2015; ISBN 978-1-107-09264-8.
3. Keyes, A.K.M.; Crawford-Brown, D. The Changing Influences on Commuting Mode Choice in Urban England under Peak Car: A Discrete Choice Modelling Approach. *Transp. Res. Part F Traffic Psychol. Behav.* **2018**, *58*, 167–176. [CrossRef]
4. Karlaftis, M.G.; Vlahogianni, E.I. Statistical Methods versus Neural Networks in Transportation Research: Differences, Similarities and Some Insights. *Transp. Res. Part C Emerg. Technol.* **2011**, *19*, 387–399. [CrossRef]
5. Martín-Baos, J.Á.; López-Gómez, J.A.; Rodriguez-Benitez, L.; Hillel, T.; García-Ródenas, R. A Prediction and Behavioural Analysis of Machine Learning Methods for Modelling Travel Mode Choice. *Transp. Res. Part C Emerg. Technol.* **2023**, *156*, 104318. [CrossRef]
6. Chang, X.; Wu, J.; Liu, H.; Yan, X.; Sun, H.; Qu, Y. Travel Mode Choice: A Data Fusion Model Using Machine Learning Methods and Evidence from Travel Diary Survey Data. *Transp. A Transp. Sci.* **2019**, *15*, 1587–1612. [CrossRef]
7. Deng, Y. Application of Machine Learning with a Surrogate Model to Explore Seniors' Daily Activity Patterns. *Transp. Lett.* **2022**, *14*, 972–982. [CrossRef]

8. Hillel, T.; Bierlaire, M.; Elshafie, M.Z.E.B.; Jin, Y. A Systematic Review of Machine Learning Classification Methodologies for Modelling Passenger Mode Choice. *J. Choice Model.* **2021**, *38*, 100221. [CrossRef]
9. Khalid, B.; ur Rehman, Z.; Haider, F.; Khan, A.H.; Hashmi, Q.N.; Raza, A.; Jameel, M.S. Regression Approach to Analyze the Travel Characteristics of University Students. *Transp. Lett.* **2024**, *1*, 1–16. [CrossRef]
10. Martín-Baos, J.Á.; García-Ródenas, R.; Rodriguez-Benitez, L. Revisiting Kernel Logistic Regression under the Random Utility Models Perspective. An Interpretable Machine-Learning Approach. *Transp. Lett.* **2021**, *13*, 151–162. [CrossRef]
11. Wang, F.; Ross, C.L. Machine Learning Travel Mode Choices: Comparing the Performance of an Extreme Gradient Boosting Model with a Multinomial Logit Model. *Transp. Res. Rec.* **2018**, *2672*, 35–45. [CrossRef]
12. Hermaputi, R.L.; Hua, C. Decoding Jakarta Women's Non-Working Travel-Mode Choice: Insights from Interpretable Machine-Learning Models. *Sustainability* **2024**, *16*, 8454. [CrossRef]
13. Xie, C.; Lu, J.; Parkany, E. Work Travel Mode Choice Modeling with Data Mining: Decision Trees and Neural Networks. *Transp. Res. Rec.* **2003**, *1854*, 50–61. [CrossRef]
14. Omrani, H. Predicting Travel Mode of Individuals by Machine Learning. *Transp. Res. Procedia* **2015**, *10*, 840–849. [CrossRef]
15. Hagenauer, J.; Helbich, M. A Comparative Study of Machine Learning Classifiers for Modeling Travel Mode Choice. *Expert Syst. Appl.* **2017**, *78*, 273–282. [CrossRef]
16. 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 Behav. Soc.* **2020**, *20*, 22–35. [CrossRef]
17. Dabiri, S.; Heaslip, K. Inferring Transportation Modes from GPS Trajectories Using a Convolutional Neural Network. *Transp. Res. Part C Emerg. Technol.* **2018**, *86*, 360–371. [CrossRef]
18. Ma, Y.; Zhang, Z. Travel Mode Choice Prediction Using Deep Neural Networks With Entity Embeddings. *IEEE Access* **2020**, *8*, 64959–64970. [CrossRef]
19. Yazdizadeh, A.; Patterson, Z.; Farooq, B. Ensemble Convolutional Neural Networks for Mode Inference in Smartphone Travel Survey. *IEEE Trans. Intell. Transport. Syst.* **2020**, *21*, 2232–2239. [CrossRef]
20. Yu, J.J.Q. Travel Mode Identification With GPS Trajectories Using Wavelet Transform and Deep Learning. *IEEE Trans. Intell. Transport. Syst.* **2021**, *22*, 1093–1103. [CrossRef]
21. Cheng, L.; Chen, X.; Wei, M.; Wu, J.; Hou, X. Modeling Mode Choice Behavior Incorporating Household and Individual Sociodemographics and Travel Attributes Based on Rough Sets Theory. *Comput. Intell. Neurosci.* **2014**, *2014*, 560919. [CrossRef] [PubMed]
22. Naseri, H.; Waygood, E.O.D.; Patterson, Z.; Alousi-Jones, M.; Wang, B. Travel Mode Choice Prediction: Developing New Techniques to Prioritize Determinant and Interpret Black-Box Machine Learning Techniques. *SSRN* **2022**, 42p. [CrossRef]
23. Tamim Kashifi, M.; Jamal, A.; Samim Kashefi, M.; Almoshaogeh, M.; Masiur Rahman, S. Predicting the Travel Mode Choice with Interpretable Machine Learning Techniques: A Comparative Study. *Travel Behav. Soc.* **2022**, *29*, 279–296. [CrossRef]
24. Zhang, J.; Feng, T.; Timmermans, H.; Lin, Z. Improved Imputation of Rule Sets in Class Association Rule Modeling: Application to Transportation Mode Choice. *Transportation* **2023**, *50*, 63–106. [CrossRef]
25. García-García, J.C.; García-Ródenas, R.; López-Gómez, J.A.; Martín-Baos, J.Á. A Comparative Study of Machine Learning, Deep Neural Networks and Random Utility Maximization Models for Travel Mode Choice Modelling. *Transp. Res. Procedia* **2022**, *62*, 374–382. [CrossRef]
26. Bierlaire, M. Swissmetro. Available online: http://transp-or.epfl.ch/documents/technicalReports/CS_SwissmetroDescription.pdf (accessed on 5 January 2025).
27. Arkoudi, I.; Krueger, R.; Azevedo, C.L.; Pereira, F.C. Combining Discrete Choice Models and Neural Networks through Embeddings: Formulation, Interpretability and Performance. *Transp. Res. Part B Methodol.* **2023**, *175*, 102783. [CrossRef]
28. Han, Y.; Pereira, F.C.; Ben-Akiva, M.; Zegras, C. A Neural-Embedded Discrete Choice Model: Learning Taste Representation with Strengthened Interpretability. *Transp. Res. Part B Methodol.* **2022**, *163*, 166–186. [CrossRef]
29. Mabit, S.L. Empirical Analyses of a Choice Model That Captures Ordering among Attribute Values. *J. Choice Model.* **2017**, *25*, 3–10. [CrossRef]
30. Nam, D.; Cho, J. Deep Neural Network Design for Modeling Individual-Level Travel Mode Choice Behavior. *Sustainability* **2020**, *12*, 7481. [CrossRef]
31. Sifringer, B.; Lurkin, V.; Alahi, A. Enhancing Discrete Choice Models with Representation Learning. *Transp. Res. Part B Methodol.* **2020**, *140*, 236–261. [CrossRef]
32. Zhang, H.; Feng, L.; Zhang, X.; Yang, Y.; Li, J. Necessary Conditions for Convergence of CNNs and Initialization of Convolution Kernels. *Digit. Signal Process.* **2022**, *123*, 103397. [CrossRef]