eISSN 2783-6851

NEW TRENDS in COMPUTER SCIENCES

2025
Volume 3
Issue 1
Pages 64–82
https://doi.org/10.3846/ntcs.2025.23932

ENHANCING PREDICTION OF RIDE-HAILING FARES USING ADVANCED DEEP LEARNING TECHNIQUES

Tuyet Ngoc Thi HUYNH[1], Huu Dat BUI[1], Tuyet Nam Thi NGUYEN [2], Tan Dat TRINH [1✉]

[1]*Faculty of Information Science, Saigon University, Ho Chi Minh City, Vietnam*
[2]*Faculty of Engineering and Technology, Saigon University, Ho Chi Minh City, Vietnam*

Article History:
- received 14 May 2025
- accepted 18 July 2025

Abstract. Fare prediction is a critical component of online ride-hailing services, as it significantly influences consumer decision-making and enhances operational efficiency for service providers. Reliable fare prediction is especially important in dynamic pricing environments, where fares are affected by factors such as demand fluctuations, traffic conditions, and weather patterns. This study aims to enhance fare prediction in ride-hailing services by utilizing advanced deep learning models. Using a comprehensive dataset of Uber and Lyft fare data collected in Boston during the winter of 2018, we evaluated three deep learning architectures: Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and BiLSTM with an attention mechanism (BiLSTM + Attention). The results showed that the BiLSTM + Attention model achieved the highest prediction accuracy, making it the most effective approach for fare prediction. However, its longer training time poses limitations for time-sensitive applications. Conversely, the LSTM model provided a strong balance between predictive accuracy and computational efficiency, making it a suitable alternative for scenarios that require faster model deployment. Additionally, our analysis identified key factors influencing fare variability – such as trip distance, time of day, and weather conditions – highlighting the importance of feature selection in enhancing model performance. By improving fare prediction accuracy, this study offers valuable insights for optimizing dynamic pricing strategies, enhancing consumer satisfaction, and helping ride-hailing platforms better manage supply–demand imbalances. These findings provide a foundation for future research exploring hybrid models and real-time data integration to further improve predictive capabilities in ride-hailing services.

Keywords: fare prediction, ride-hailing, LSTM, BiLSTM, Attention mechanism.

✉Corresponding author. E-mail: ***trinhtandat@sgu.edu.vn***

1. Introduction

Fare pricing is an important factor influencing consumer decisions in online ride-hailing services. Pricing strategies directly impact customer demand, company revenue, and market competition. Various dynamic factors contribute to fluctuations in ride fares, including weather conditions, peak-hour demand, city events, and unforeseen crises such as pandemics (Liu et al. 2021; Rangel et al., 2022). Thus, a robust and accurate fare prediction model is essential for both ride-sharing companies and consumers, ensuring transparency and optimized pricing. Ride-hailing services like Uber and Lyft have transformed the transportation industry by positioning themselves as cost-effective and accessible alternatives to traditional taxis. In major metropolitan areas such as Boston, these services advertise base fares ranging from $2 to $8 per mile, excluding luxury vehicle segments (Dogo et al., 2020; Haolun Huang, 2023; Schwieterman, 2019). However, the actual cost incurred by a passenger is influenced by multiple dynamic factors beyond the base fare. Price surges, which can cause the fare to increase between 1.5 to 3 times the standard rate, contribute to significant price volatility, creating challenges for both users and regulators (Battifarano & Qian, 2019; Farris et al., 2014).

A key concern for consumers is the unpredictability of pricing, which can undermine trust in ride-hailing applications. Unforeseen fare hikes during peak hours, adverse weather conditions, or major public events add complexity to fare structures. While this price variability may seem unfair from a consumer perspective, it functions as a dynamic pricing strategy that helps businesses balance supply and demand efficiently. Previous studies have shown that surge pricing mechanisms enable ride-hailing companies to optimize fleet utilization, manage driver availability, and maximize revenue while maintaining service efficiency (Ashkrof et al., 2022; Battifarano & Qian, 2019; Castillo et al., 2017; Garg & Nazerzadeh, 2020). Furthermore, several studies have examined ride-hailing fare determinants, revealing that time of day, weather conditions, trip distance, ride demand, and local events significantly impact fare variability (Rangel et al., 2022). Additionally, research has indicated that regulatory policies and fuel prices play an essential role in price fluctuations (Battifarano & Qian, 2019; Sriwongphanawes & Fukuda, 2024).

Given these challenges, this study aims to explore the key factors influencing Uber and Lyft fares. Moreover, to enhance the accuracy of fare prediction, we propose the implementation of deep learning techniques, specifically Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and BiLSTM with Attention mechanisms (BiLSTM + Attention). These models are particularly effective in processing time-series data and can capture complex non-linear relationships in pricing trends without requiring extensive manual feature engineering (Eslami & Ghaderi, 2024; S. Li et al., 2024). By training these models on datasets that include variables such as time, ambient temperature, trip origin, destination, and external conditions, we aim to improve fare prediction accuracy over traditional statistical and machine learning approaches. Our research seeks to provide deeper insights into ride-hailing pricing strategies and to develop an advanced predictive framework for dynamic fare estimation. The findings will not only benefit consumers by offering better price transparency but also assist ride-hailing companies in optimizing pricing algorithms for improved efficiency and fairness.

2. Related works

Numerous studies have explored the application of machine learning and deep learning techniques to enhance the predictability of Uber and Lyft fares. These studies demonstrate the effectiveness of various predictive models in capturing dynamic pricing patterns influenced by factors such as trip distance, time of day, demand-supply imbalances, and external conditions. Regarding the machine learning-based approaches, traditional machine learning models have been extensively used to forecast Uber and Lyft fares due to their efficiency and interpretability. For instance, Battifarano and Qian (2019) employed a Linear Regression-based model enhanced with L1 regularization and pattern clustering techniques. Their model successfully predicted Uber's surge multiplier in Pittsburgh up to two hours in advance and Lyft's surge multiplier up to 20 minutes in advance, outperforming alternative non-linear models in various locations (Battifarano & Qian, 2019).

Random Forest (RF) models have also demonstrated robust performance in fare prediction tasks. For instance, Silveira-Santos et al. (2023) predicted Uber fares in Madrid using

Random Forest, Decision Tree, as well as Linear Regression models. Their results revealed that Random Forest had better accuracy and effectively identified discrepancies caused by peak-hour surges. Similarly, Hao Huang (2023) introduced a Random Forest-based system for predicting Uber fares using data from various locations and highlighted its superior predictive performance over baseline models. Moreover, other machine learning approaches such as XGBoost have demonstrated success in capturing nonlinear fare trends and mitigating overfitting issues. XGBoost regression model was used for predicting travel times in ride-hailing services with improved accuracy by identifying inlier and extreme-conditioned trips (Kankanamge et al., 2019). XGBoost was also mentioned to outperform Random Forest model and was highlighted to be suitable for high- accuracy fare prediction (Jashwanth et al., 2024; Kankanamge et al., 2019; Poongodi et al., 2022). Moreover, Y. Chen et al. (2015) used various machine learning algorithms, including AdaBoost, Gradient Boosting, K-Nearest Neighbors (KNN), and Bagging with extra tree classifiers, to predict Uber fare and mentioned that XG-Boost consistently demonstrated superior predictive performance.

Deep learning models have gained increasing attention for predicting ride-hailing fares due to their ability to model complex and non-linear relationships in time-series data. Long Short-Term Memory (LSTM) networks have been particularly effective in this domain due to their ability to capture long-term dependencies in sequential data (Guo et al., 2024; Kumar et al., 2021). LSTM model was also used to predict Uber travel demand and exhibited good performance (Alghamdi et al., 2022). Additionally, Bi-directional LSTM (BiLSTM) model has shown improved performance by processing input data in both forward and backward directions, offering a more comprehensive understanding of temporal dependencies. X. Zhao et al. (2023) proposed an Random Forest-Bidirectional LSTM (RF-BiLSTM) model integrated with an attention mechanism, which significantly enhanced prediction precision for Uber pickup data in New York City. The model demonstrated improved forecasting accuracy compared to standard LSTM models by selectively focusing on key temporal features. Furthermore, Z. Li et al. (2024) presented a model combining BiLSTM with an attention mechanism to improve the prediction of dynamic fare fluctuations and short-term traffic flow. Their proposed model outperformed traditional LSTM models and other approaches like Autoregressive Integrated Moving Average (ARIMA) and Support Vector Regression (SVR). X. Zhao et al. (2023) combined the attention mechanism and the RF-BiLSTM as well as XGboost-BiLSTM models to enhance the prediction accuracy of Uber pickups and these proposed models outperformed traditional LSTM model. CNN-LSTM hybrid models have also emerged as powerful tools for fare prediction. For instance, Ara and Hashemi (2022) proposed a CNN-LSTM-Autoencoder model to predict ride-hailing demand patterns in New York City, demonstrating improved predictive accuracy by capturing both spatial and temporal dependencies.

Recent studies emphasize the importance of incorporating external factors such as weather conditions, traffic congestion, and local events to improve fare prediction models. For instance, weather conditions such as rainfall and strong winds significantly influenced ride demand, impacting dynamic fare pricing (Liu et al., 2021). Furthermore, Silveira-Santos et al. (2023) examined the impact of demand-supply imbalances on Uber fares in Madrid, demonstrating how city events, holidays, and unexpected disruptions contribute to fare volatility. Moreover, the balance between the number of drivers (supply) and passenger requests (demand) can be the primary driver of dynamic pricing. Particularly, during peak hours or special events, demand often outstrips supply and lead to higher fares (Karamanis et al., 2021). Fares tend also to be higher during rush hours, late nights, and weekends., thus, predictive models must account for these temporal patterns (Battifarano & Qian, 2019; Sindhu et al., 2022; Zhang et al., 2017). Previous studies have also investigated the impact of unforeseen crises, such as the COVID-19 pandemic, on fare prediction models. The ride-hailing fares decreased significantly in 2020 due to reduced demand during COVID-19 (Silveira-Santos et al., 2024). Moreover, another study suggest that COVID-19-related fears influenced driver behavior, resulting in a lower acceptance rate for ride requests, which could have indirectly impacted fare dynamics (Ashkrof et al., 2022). Table 1 provides a summary of several previous studies focused on predicting ride-hailing fares and demand.

Table 1. Summary of several previous studies on using machine learning and deep learning for predicting ride-hailing fares and demands.

Target	Model	Evaluation metrics	References
Uber travel demand	Multilayer perception	RMSE: 0.1446; SMAPE: 0.1147	Alghamdi et al. (2022)
	Multi-LSTM	RMSE: 0.1381; SMAPE: 0.1074	
	Sequence to Sequence	RMSE: 0.1133; SMAPE: 0.0965	
	MSP-TCN	RMSE: 0.0971; SMAPE: 0.0793	
UberX fare	Linear Regression	RMSE: 3.41, MAPE: 8.01	Silveira-Santos et al. (2023)
	Decision Tree	RMSE: 3.85, MAPE: 6.60	
	Random Forest	RMSE: 3.40, MAPE: 6.32	
Uber fare price	Linear Regression	RMSE: 1.718	Hao Huang (2023)
	Decision Tree	RMSE: 1.277	
	Random Forest	RMSE: 1.264	
Uber pickups	Attention-RF-BiLSTM	MAE: 0.0283, MSE: 0.0015	X. Zhao et al. (2023)
	Attention-XGBoost-BiLSTM	MAE: 0.0306, MSE: 0.0018	
	LSTM	MAE: 0.0346, MSE: 0.0025	
Uber passenger demand	LSTM	RMSE:17.16, MAE: 11.46	J. Zhao et al. (2023)
	ST-ResNet	RMSE:14.85, MAE: 8.87	
	Unified Spatio-Temporal Network	RMSE:14.22, MAE: 8.39	
	STIMN	RMSE:14.04, MAE: 8.32	
Ride-hailing services	LSTM	SMAPE: 7.53, RMSE: 179.21	L. Chen et al. (2021)
	CNN-LSTM	SMAPE: 7.48, RMSE: 179.82	
	CNN	SMAPE: 7.79, RMSE: 183.03	
	UberNet	SMAPE: 7.31, RMSE: 177.84	
Taxi fare prices	Linear Regression	RMSE: 0.32	Chou et al. (2023)
	Random Forest	RMSE: 0.59	
	Multilayer perception	RMSE: 0.25	
	LSTM	RMSE: 0.098	

Note: MSP-TCN – Multi-stage Probabilistic Temporal Convolution Network; STIMN – Spatio-Temporal Information Modulation Network; ATT-BiLSTM – Fusion Attention Mechanism Bidirectional Long Short-Term Memory; RMSE – Root mean square error; SMAPE – Symmetric mean absolute percentage error; MSE – Mean squared error; MAE –: Mean absolute error.

3. Dataset description

We utilize the "Uber and Lyft Cab Prices" dataset – available on Kaggle (https://www.kaggle.com/datasets/ravi72munde/uber-lyft-cab-prices/data) – to investigate the factors affecting ride fares and implement deep learning models for fare prediction. The dataset comprises two components: "Cab_ride" and "Weather". The "Cab_ride" dataset was collected every five minutes and contained 693,071 records with 10 features, such as distance, cab type, timestamp, destination, source, ride prices, cab name, surge multiplier, and record ID. The "Weather" dataset was recorded hourly and had 6,276 records with 8 features, including temperature, location, cloud cover, pressure, humidity, wind index, and timestamp. This data provides comprehensive insights into ride characteristics and environmental conditions, supporting robust predictive modeling with deep-learning techniques.

For data preprocessing, the timestamp values in both datasets were transformed into the "dd/mm/yy hh:mm:ss" format to maintain a standardized temporal structure. Next, the Uber and Lyft vehicle models were grouped into six categories: "Share" ([UberPool, Shared]), "Normal" ([UberX, Lyft]), "SUV" ([UberXL, Lyft XL]), "LUX" ([Black, Lux Black]), "LUX SUV" ([Black SUV, Lux Black XL]), and "Other" ([Lux, Taxi, WAV]). Then, rows with missing "price" values in the "Car_ride" dataset were removed to ensure data integrity, while missing "rain" values in the "Weather" dataset were filled with zero. Additionally, the "humidity" column was rescaled from 0–1 to 1–100 for consistency. The two datasets were then merged into a new one to enable comprehensive analysis and improved predictive modeling.

After preprocessing, the datasets are free of missing value. The statistical description of the merged dataset, which was used for further analysis and fare prediction, is shown in Table 2. As shown in this table, the data likely came from cold months (November to December) of 2018. Rides were likely to concentrate in morning to early evening, and the surge of fare prices would depend on trip distance or vehicle models. Additionally, weather conditions showed relatively low ambient air temperature, high humidity, high cloudy skies, and low rainfall. These meteorological conditions would represent a cold urban environment.

Table 2. Statistical description of the merged dataset.

Variable	Unit	Range	Q1	Q2	Q3	Mean	SD
Day	–	1–30	16	27	29	22.02	9.34
Month	–	11–12	11	11	12	11.35	0.48
Year	–	2018	2018	2018	2018	2018	0
Hour	–	0–23	6	10	17	11.12	6.57
Distance	–	0.02–7.86	1.28	2.15	2.94	2.19	1.14
Price	USD	2.5–97.5	9	13.5	22.5	16.54	9.32
Surge multiplier	–	1–3	1	1	1	1.02	0.1
Temperature	°F	19.62–55.41	37.28	39.63	42.74	39.28	5.48
Cloud cover	–	0–1	0.46	0.76	0.97	0.68	0.3
Pressure	mb	988.25–1035.12	994.1	1000.85	1014.57	1005.02	12.71
Rain	inches	0–0.78	0	0	0	0.01	0.05
Humidity	%	45–99	66	72	87	75.3	11.96
Wind	mph	0.29–18.18	4.99	8.88	10.25	7.72	3.44

Note: Q1 – the 25th percentile; Q2 – the 50th percentile; Q3 – the 75th percentile; SD – Standard deviation. Several variables which are non-numeric and likely used for grouping or classification include vehicle models, destination, source, product ID, and location.

4. Proposed method

To predict Uber and Lyft fares, we employed three deep learning models: LSTM, BiLSTM, and BiLSTM combined with an Attention mechanism (BiLSTM + Attention). The architectures of the LSTM and BiLSTM models are illustrated in Figure 1.

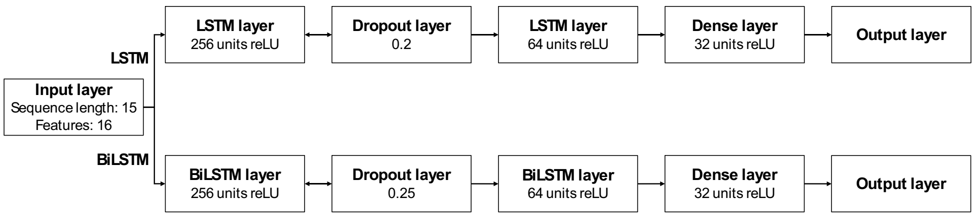


Figure 1. The architecture of LSTM and BiLSTM models

Our proposed models were trained on an extensive dataset comprising trip details and meteorological conditions in Boston, USA during winter of 2018. We fed input features with a size of 16 and a sequence length of 15 into the LSTM and BiLSTM models for learning context information and capturing long-term dependency, which frequently appears in time-series. These models are developments and modifications of Recurrent Neural Networks (RNNs) (Z. Li et al., 2024). The LSTM model (Vuong et al., 2025) is a more advanced version of RNN designed to mitigate the vanishing gradient problem. In addition, LSTM model mitigates the vanishing gradient problem by incorporating a cell state component per timestep t. The memory block receives three essential parts: the cell state C_{t-1}, the hidden state h_{t-1}, which are from the previous timestep and current timestep x_t. The cell state plays an important role as long – term memory, while hidden state is short – term memory. In addition, the memory block has three gates: forget gate, input gate and output gate to maintain important information. Information stored in this component can be added or deleted using these gates. Figure 2 shows an example of LSTM unit.

An advanced type of LSTM model, the BiLSTM model (Ihianle et al., 2020; X. Zhao et al., 2023) is a deep learning model that is often used for time series data analysis. Unlike LSTM, the Bidirectional LSTM (BiLSTM) can process data in both forward and backward directions, thereby enhancing its performance capabilities. This bidirectional nature enables the model to capture dependencies from both past and future contexts, improving its ability to understand relationships within the data. As a result, BiLSTM often outperforms standard LSTM models, which can only process information in a single direction. However, these advantages come at a cost: the construction and training of BiLSTM models are typically more resource-intensive, often requiring nearly twice the computational cost of LSTM. Furthermore, when working with large datasets, BiLSTM models tend to have significantly longer training times. In applications that demand real-time data processing, these limitations may hinder the practicality of using BiLSTM.

Figure 3 shows an example of the BiLSTM model. As mentioned previously, there are forward and backward layers in a bidirectional layer. These sub-layers use many LSTM models to train data., however, the forward layer is responsible for training data in the direction from the beginning of the time series to the end of the time series, and the backward layer is responsible for training data in the direction from the end of the time series to the beginning of it.

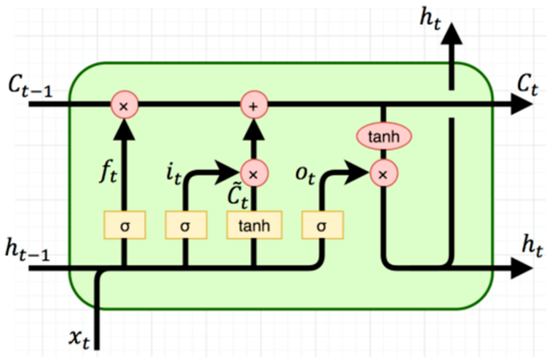


Figure 2. A memory block in the LSTM model (Vuong et al., 2022, 2025)

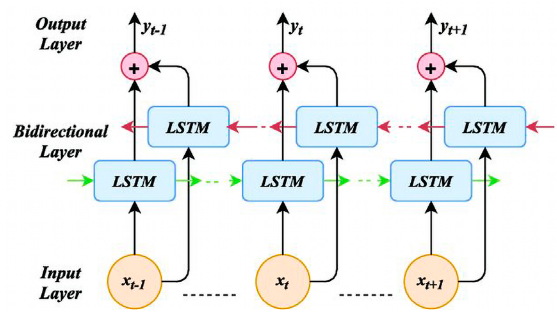


Figure 3. The BiLSTM model's architecture (Ihianle et al., 2020)

4.1. Proposed BiLSTM + Attention model

To effectively model the complexity of fare data, we introduce a hybrid deep learning architecture that combines BiLSTM with an Attention mechanism (hereafter BiLSTM + Attention) (Z. Li et al., 2024; Pham et al., 2024; Vuong et al., 2025). This design enables the model to focus selectively on the most relevant parts of the input sequence, thereby improving the accuracy of fare price predictions. The Attention mechanism assigns dynamic weights to each time step in the input sequence, enabling the model to emphasize the most influential time steps when making predictions. It produces a context vector – a weighted summary that captures the most relevant sequence information – thereby enhancing the model's decision-making process.

Figure 4 illustrates an example of the BiLSTM+Attention model. As shown in the figure, the input, output, forward, and backward layers function similarly to those in the standard BiLSTM model. However, the key distinction lies in the incorporation of the attention mechanism. This mechanism aims to generate an attention vector that highlights the most relevant features the model should focus on when making predictions.

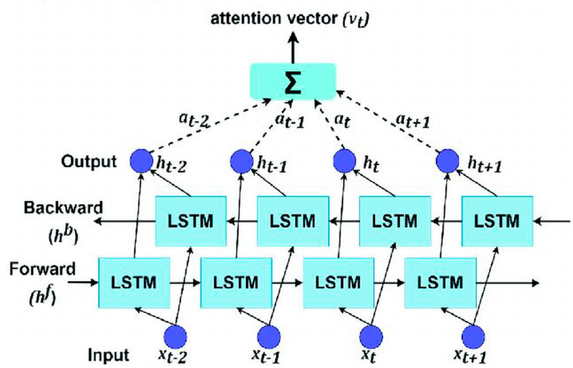


Figure 4. The BiLSTM – attention model's architecture (Yousaf & Nawaz, 2022)

Our hybrid BiLSTM + Attention model effectively captures the complementary strengths of recurrent neural networks and context-aware feature weighting. The BiLSTM component plays a critical role in modeling fare price data by capturing both past and future dependencies, ensuring temporal continuity while maintaining model complexity. Meanwhile, the Attention mechanism enables the model to dynamically prioritize significant time steps, allowing it to better adapt to irregular patterns in fare fluctuations. Combined, these components work synergistically to enhance predictive performance, with both the BiLSTM and Attention mechanisms contributing significantly to the model's effectiveness.

4.2. Training process

The dataset was divided into three subsets: 70% for training (approximately 411,394 records for Uber and 388,804 records for Lyft), 15% for validation (approximately 89,656 records for Uber and 83,315 records for Lyft), and the remaining 15% for testing, with the test set containing the same number of records as the validation set for both services. Our Validation was conducted using a time-series hold-out strategy. Due to the sequential nature of fare data, we split the dataset in chronological order. This approach helps prevent data leakage and simulates a real-world forecasting setting where future data is inaccessible during training. For model training, we conducted 100 epochs with a batch size of 128 for the LSTM model, while the BiLSTM and BiLSTM + Attention models were trained with a batch size of 256. We utilized the Adam optimizer, a widely used adaptive learning rate optimization algorithm known for its computational efficiency and low memory requirements. Mean Squared Error (MSE) was selected as the loss function to measure prediction error, ensuring the models effectively minimized the variance between predicted and actual fare values.

During the training process, we observed that deep learning models often face challenges such as slower convergence rates and a heightened risk of overfitting when applied to large datasets. To mitigate these issues, we implemented an early stopping mechanism, which halts the training process if no improvement is observed after 10 consecutive epochs. This approach effectively prevents overfitting and optimizes model performance by ensuring the models generalize well to unseen data.

5. Results and discussion

5.1. Factors influencing fares and demand in Uber and Lyft services

5.1.1. An analysis of the demand for Uber and Lyft services

To assess the demand for Uber and Lyft services, we analyzed the "Cab_ride" dataset. As shown in Figure 5, 51.82% of trips (330,568 rides) were taken with Uber, while 48.18% (307,408 rides) were with Lyft, indicating a slight consumer preference for Uber. Although the difference in market share was relatively small, it suggested a marginal advantage in favour of Uber. Several factors might account for this discrepancy. For example, Uber's international presence likely enhanced its brand recognition and customer reach. Additionally, Uber's earlier entry into the market may have provided a competitive advantage in terms of consumer trust, familiarity, and market penetration, contributing to its slight dominance over Lyft.

Figure 5. The demand for services measured by the number of rides

5.1.2. An analysis of service demand based on customers' preferred car types

Figure 6 shows the frequently selected car types in Uber and Lyft services during November and December of 2018. The data indicated that UberXL, WAV (Wheelchair Accessible Vehicle), and Black SUVs accounted for 55,096 rides, all belonging to Uber. Their popularity is linked to capacity and accessibility. Particularly, UberXL and Black SUVs offered seven seats for comfortable group travel, while WAV was designed for wheelchair users with enhanced safety features.

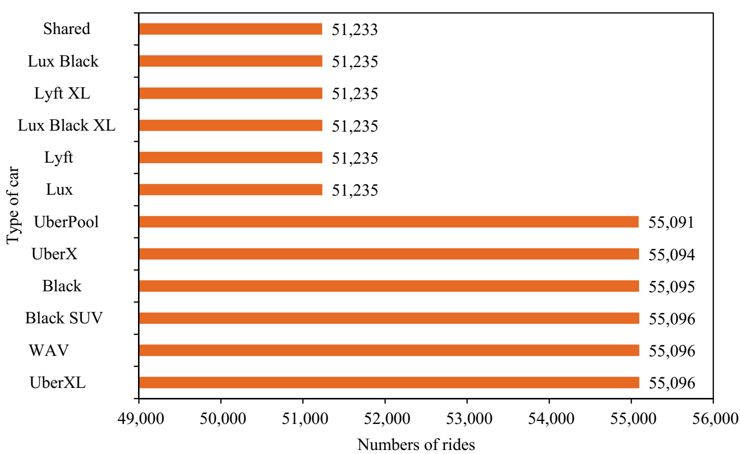
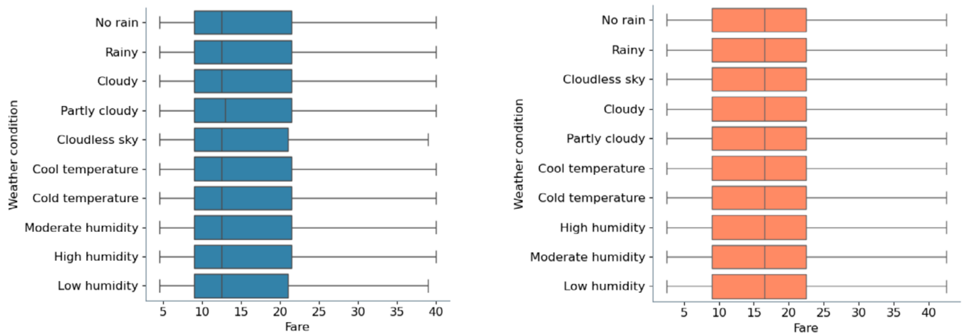


Figure 6. Car types with the highest preference

5.1.3. Influence of weather conditions on trip fares

Figure 7 illustrates the distribution of Uber and Lyft fares across various weather conditions. The data suggested that weather conditions have a minimal impact on overall fare structures for both services because there were no significant differences in fares among weather conditions, such as rain, cloudiness, and humidity levels. Moreover, both Uber and Lyft exhibited similar fare distribution patterns, with most fares falling within the \$10 to \$25 range.

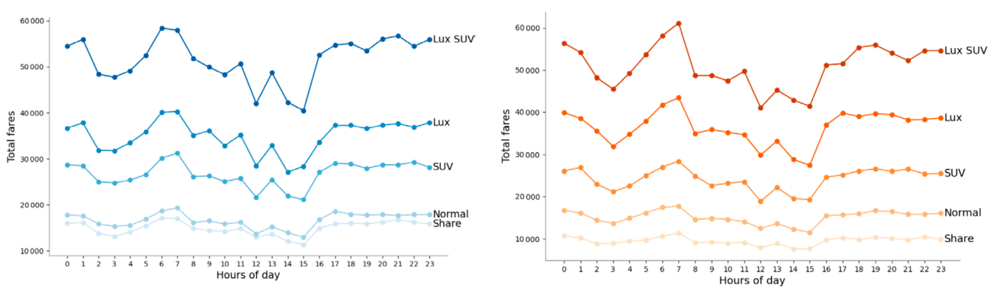


a) **b)**

Figure 7. Fares of Uber (a) and Lyft (b) in several weather conditions

5.1.4. Influence of time on fares

Figure 8 illustrates the hourly variation in fares across different vehicle models for Uber and Lyft. To ensure a more accurate representation of temporal fare fluctuations, ride data for Saturday and Sunday were excluded from the analysis, as weekends often deviate from typical weekday patterns. In the lower-priced segments, such as the “Normal” and “Share” vehicle types, fares remain relatively stable throughout the day. In contrast, higher-priced segments, including “SUV,” “Lux,” and “Lux SUV,” exhibit significant price variability. Both Uber and Lyft show a similar trend, with fares starting to rise around 3 AM, peaking between 6 and 7 AM, followed by a sharp decline until approximately 3 PM, after which fares increase again during the evening hours. The fares for these services tend to be higher during these peak travel periods, reflecting increased demand and traffic congestion.



a) **b)**

Figure 8. Hourly variation in fares for Uber (a) and Lyft (b)

5.1.5. Influence of distance on fares

Figure 9 illustrates the relationship between fares and the distance between the source and destination for Uber and Lyft services. The positive correlation between distance and fare is observed for both Uber and Lyft. In addition, Lyft fares are more densely clustered at lower price points, indicating a stronger presence in budget-friendly, short-distance rides, particularly for trips between 0.5 to 6.5 miles. In contrast, Uber fares exhibit greater variability, especially for longer-distance trips, with Uber's distance range appearing comparatively broader. For short trips of approximately 0.5 to 3 miles, Lyft tends to offer more cost-effective options. Conversely, Uber's relatively stable pricing structure suggests that for longer trips, the fare difference may remain more substantial. Overall, Uber appears to dominate the long-distance, while Lyft is more competitive in short-distance travel.

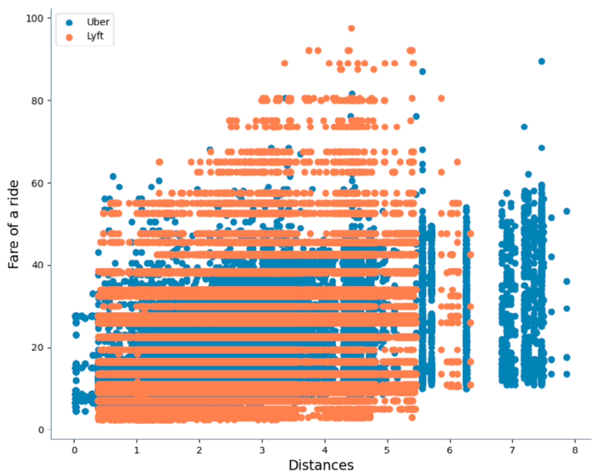
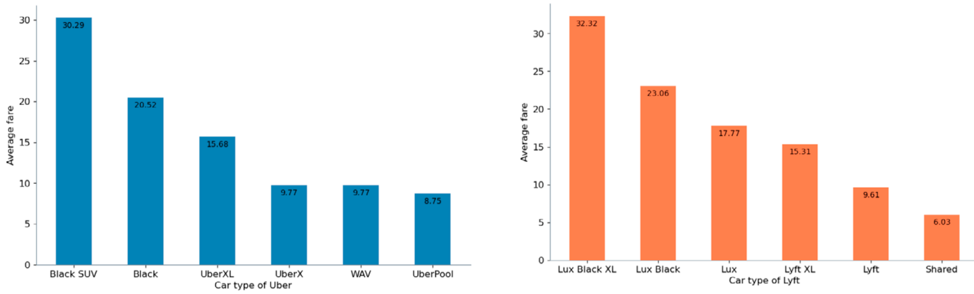


Figure 9. Scatter plot of distance and fares of Uber and Lyft

5.1.6. Influence of car's types on fares

The average fare per ride across different vehicle types of Uber and Lyft is illustrated in Figure 10. The data reveals that vehicle type is a significant factor influencing trip fares, as there is a clear distinction in fares across various vehicle types for both companies. In the premium segment, Uber's "Black SUV" and "Black" services have average fares of \$30.29 and \$20.52, respectively, while Lyft's "Lux Black XL" and "Lux Black" services are priced higher at \$32.32 and \$23.06. This shows that Lyft's premium services tend to be more expensive than Uber's by approximately \$2 to \$3, representing around a 6.7% difference.

Conversely, in the budget segment, Lyft demonstrates a significant advantage. Lyft's "Shared" service averages \$6.03, considerably lower than Uber's "UberPool" at \$8.75, reflecting a 45.11% cost reduction. This suggests that Lyft holds a competitive edge in the low-cost market, making it a preferable choice for budget-conscious customers seeking short-distance, affordable travel. The fare differences reflect strategic pricing employed by Uber and Lyft to cater to diverse customer needs and budgets. By offering distinct pricing tiers, both companies enhance customer choice while optimizing revenue streams and expanding their reach across multiple market segments.



a) **b)**

Figure 10. Average fare of Uber (a) and Lyft (b) car type

5.2. Prediction of ride-hailing fares

5.2.1. Model performance

The fares of Uber and Lyft were predicted using three deep learning models: LSTM (L. Chen et al., 2021; Chou et al., 2023), BiLSTM (X. Zhao et al., 2023), and BiLSTM combined with an attention mechanism. The performance of these models was assessed using three evaluation metrics: mean squared error (MSE), mean absolute error (MAE), and the coefficient of determination (R^2). Additionally, this study examined the learning behavior and stability of the models, considering factors such as learning rate, risks of overfitting and underfitting, and the models' ability to generalize and predict accurately on new data. The training process was conducted using two T4 GPUs on the Kaggle platform; without GPU acceleration, the training time would have been significantly longer.

Table 3 presents the evaluation metrics (i.e., MAE, MSE, and R^2) for three deep learning models (i.e., LSTM, BiLSTM, and BiLSTM+Attention) used to predict Uber and Lyft fares. These metrics were calculated for both training and testing datasets. Our results showed

Table 3. Evaluation metric comparison of LSTM, BiLSTM, and BiLSTM + Attention models used in this study

Data/Model	MAE		MSE		R^2	
	Train	Test	Train	Test	Train	Test
LSTM model						
Uber	1.2075	1.2207	3.6777	3.8316	0.9498	0.9475
Lyft	1.0872	1.1232	2.2263	2.4633	0.9778	0.9755
BiLSTM model						
Uber	1.1914	1.1893	3.6022	3.8003	0.9512	0.9479
Lyft	1.0582	1.0978	2.3506	2.3242	0.9783	0.9769
BiLSTM + Attention						
Uber	1.1814	1.1749	3.6621	3.7166	0.9499	0.9516
Lyft	1.0196	1.0149	2.2013	2.1441	0.9804	0.9796

that the BiLSTM+Attention model consistently outperformed the LSTM and BiLSTM models across all metrics, with the lower MAE and MSE values coupled with higher R^2 scores. This result highlighted the effectiveness of integrating the attention mechanism with BiLSTM for improving prediction accuracy in complex and time-series fare data, such as Uber and Lyft fares. Moreover, the prediction of Lyft fares slightly outperformed that of Uber, which may result from a combination of several factors such as lower fare variability, fewer outliers, and clearer feature correlations.

Figure 11 presents a comparison between actual and predicted fares using three models: LSTM, BiLSTM, and BiLSTM with an attention mechanism, evaluated independently on Uber and Lyft datasets. The red diagonal line represents the ideal prediction line (y = x), where predicted values align perfectly with actual values. Overall, the predictions for Uber data exhibited lower accuracy across all models compared to Lyft data, as the points representing

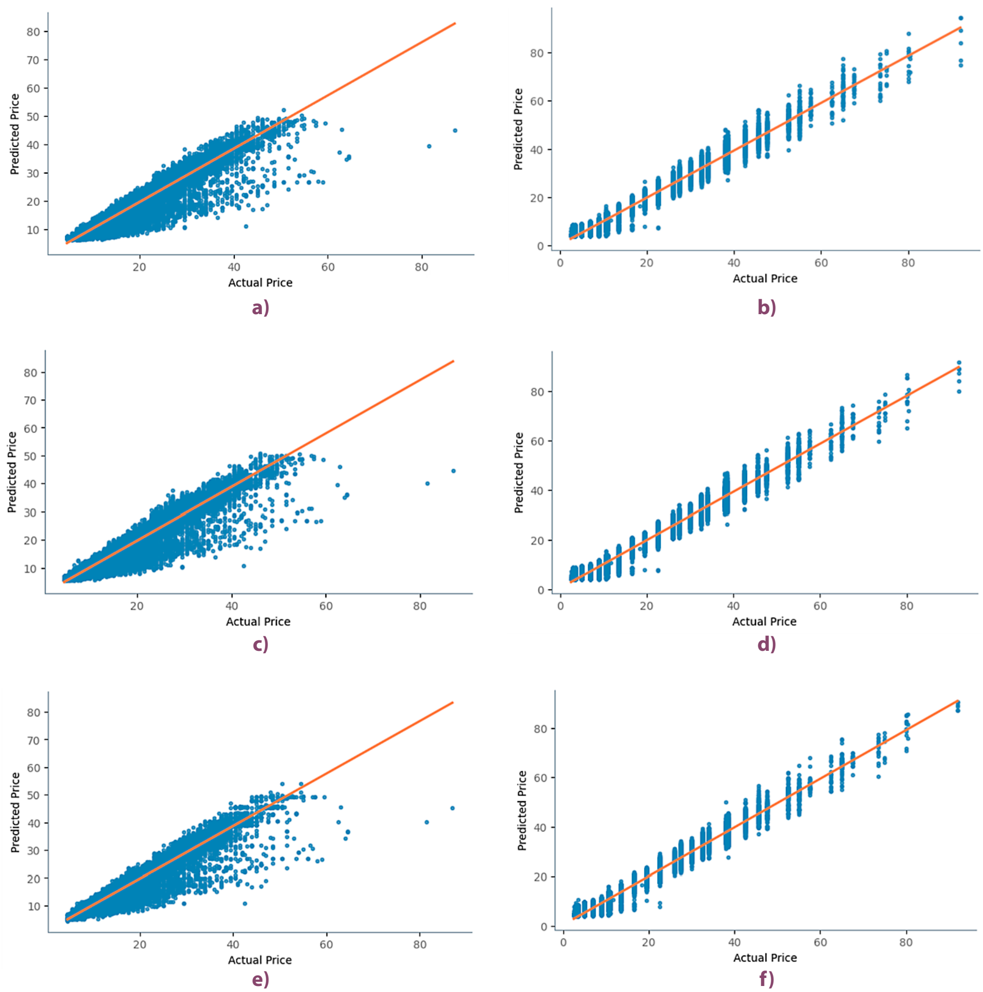


Figure 11. Scatter plots of actual and predicted fares of LSTM for Uber (a), LSTM for Lyft (b), BiLSTM for Uber (c), BiLSTM for Lyft (d), BiLSTM with attention mechanism for Uber (e), and BiLSTM with attention mechanism (f)

actual and predicted fares for Lyft data showed less deviation from the ideal prediction line. Notably, the BiLSTM with Attention model demonstrated significant performance improvements for both Uber and Lyft fare predictions. This improvement was evident in the reduced spread of predicted values around the ideal prediction line (Figure 11e and Figure 11f), indicating that the attention mechanism effectively enhanced prediction accuracy by capturing essential patterns and dependencies within the data.

Figures 12 and 13 illustrate the comparison between the training and validation performance of these models (i.e., LSTM, BiLSTM, and BiLSTM + Attention) for predicting Uber and Lyft fares. The performance of each model was evaluated based on the MAE and Loss corresponding to multiple epochs. In general, the LSTM model achieved rapid convergence with minimal fluctuations in MAE and loss values. However, the BiLSTM model demonstrated instability in its validation MAE and exhibited signs of overfitting, which was indicated by the periodic spikes (Figures 12c and 13c). The BiLSTM model with attention outperformed both

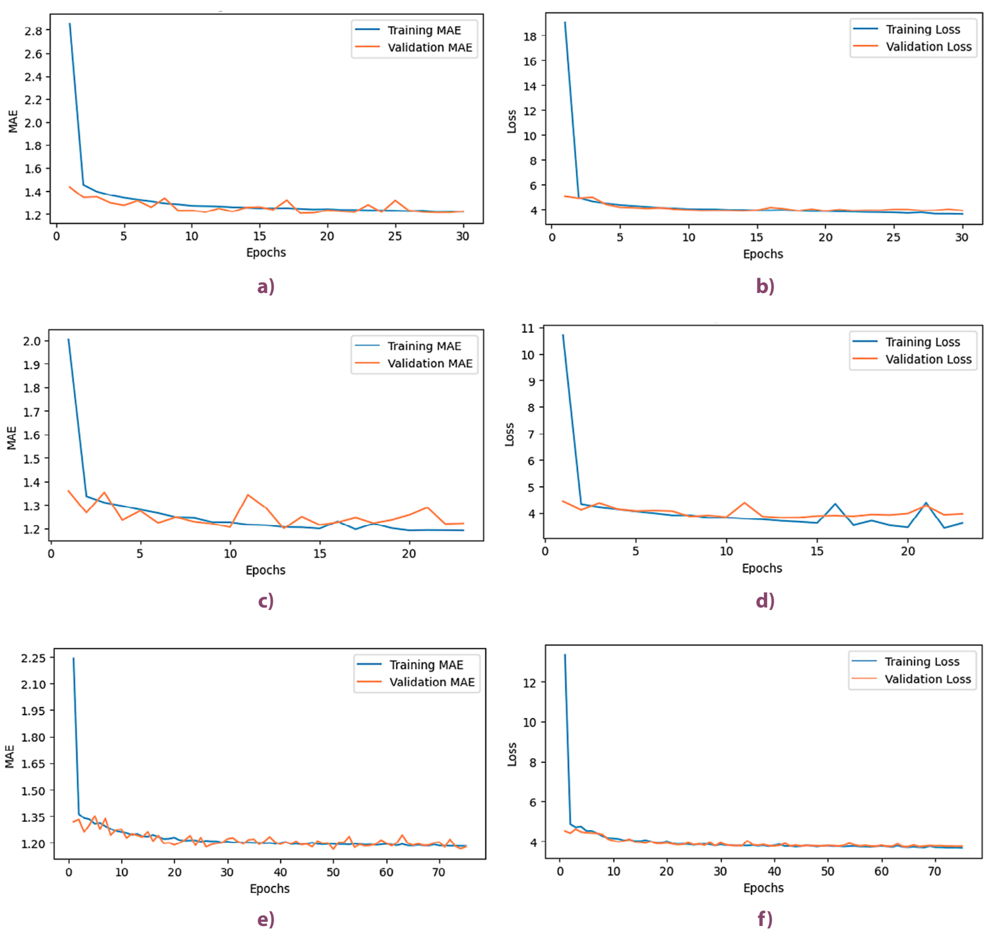


Figure 12. Training MAE, validation MAE, training loss, and validation loss of LSTM, BiLSTM, and BiLSTM + Attention mechanism models used for the prediction of Uber fares

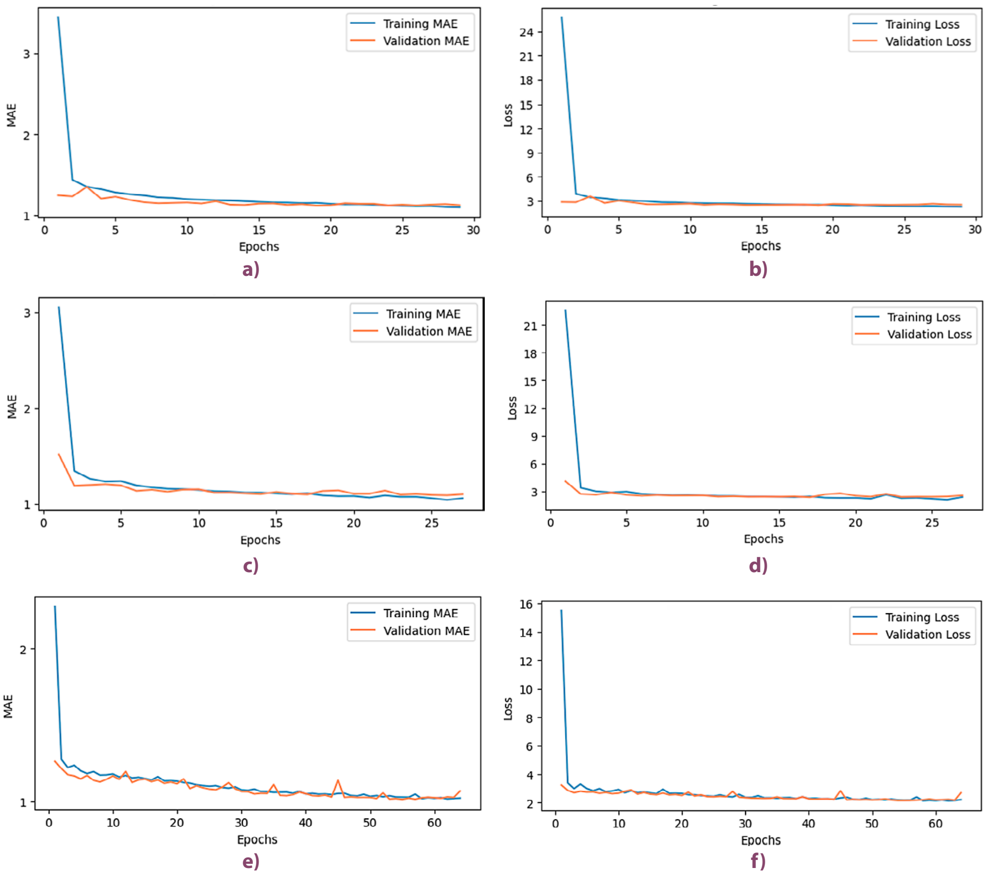


Figure 13. Training MAE, validation MAE, training loss, and validation loss of LSTM, BiLSTM, and BiLSTM + Attention mechanism models used for the prediction of Lyft fares

alternatives, achieving consistently lower MAE values and stable loss profiles, suggesting the attention mechanism would effectively enhance the model learning by prioritizing key features in the input data.

In the initial training phase (epochs 1 to 5), the LSTM model demonstrated a rapid learning rate, indicated by a significant reduction in both MAE and loss values. This rapid decline suggested that the model effectively captured core data characteristics during early training stages. However, beyond the fifth epoch, the learning rate decelerated, with the model exhibiting slight fluctuations in performance. While the LSTM model showed some degree of overfitting during the early epochs, its stability improved as training progresses.

The BiLSTM model similarly demonstrated strong initial learning capabilities, characterized by a sharp reduction in MAE and loss values during the first five epochs. However, after this point, the model exhibited noticeable instability, with an increased risk of overfitting on both Uber and Lyft datasets. This overfitting tendency was more pronounced with Uber data, likely due to the bidirectional architecture of BiLSTM, which inherently increased the model complexity and the number of trainable parameters. Given that Uber's dataset might lack sufficient richness to support this complexity, the model's susceptibility to overfitting was heightened.

The BiLSTM with Attention model demonstrated efficient learning during the initial training epochs, characterized by a rapid decline in both MAE and loss values. As training progressed, the model's performance remained stable, with evaluation metrics showing only moderate fluctuations. Notably, the model did not exhibit signs of overfitting or underfitting, suggesting strong generalization capabilities. However, this enhanced performance came at the cost of increased training time.

5.2.2. Comparison of training time of the models

Regarding training efficiency, the LSTM model required 1,193 seconds (approximately 19.88 minutes) to train on the Uber dataset and 1,071 seconds (approximately 17.85 minutes) for the Lyft dataset as shown in Table 4. These findings indicated that LSTM model was relatively efficient., however, it still required a considerable amount of time to train the model. The BiLSTM model required 1,442 seconds (approximately 24.03 minutes) for the Uber dataset and 1,579 seconds (approximately 26.32 minutes) for the Lyft dataset. The increased training time was attributed to bidirectional structure of the model, which inherently required additional computations for processing large datasets.

Table 4. Training time for the prediction of Uber and Lyft fares

	Uber fare prediction	Lyft fare prediction
LSTM	19.88 min	17.85 min
BiLSTM	24.03 min	26.31 min
BiLSTM-Attention	77.31 min (~ 1 h, 17 min, 9 s)	62.63 min (~ 1h, 2 min, 38 s)

Additionally, the BiLSTM with Attention model required 4,639 seconds (approximately 1 hour, 17 minutes, and 9 seconds) for the Uber dataset and 3758 seconds (approximately 1 hour, 2 minutes, and 38 seconds) for the Lyft dataset. The prolonged training duration of BiLSTM + Attention model (i.e., approximately 2–4 times greater than that of the other models) was largely attributed to the model's complexity, which combines bidirectional learning, attention weight computation, and evaluation of interactions between sequence time steps. Nevertheless, the BiLSTM + Attention model achieved the most stable and accurate performance, albeit with significantly longer training durations due to its intricate architecture. This trade-off would highlight the importance of model selection based on both performance needs and computational resources.

6. Conclusions

This study highlights the promising potential of deep learning models for fare prediction in ride-sharing services and emphasizes the need for further improvements to ensure practical applicability in real-world environments. We utilized a dataset containing fares for Uber and Lyft services in Boston during the winter of 2018. Our analysis identified several key factors that influence the ride-hailing fares, including distance, travel time, and vehicle types. Additionally, we provided an overview of the underlying principles of deep learning models, particularly LSTM and BiLSTM, and introduced an enhanced BiLSTM model incorporating an attention mechanism to improve predictive accuracy. Experimental results revealed that all three models achieved strong predictive performance, with R^2 values exceeding 94% for Uber and 97% for Lyft, along with MAE values slightly above 1 and MSE values slightly above 3. The BiLSTM model with attention mechanism achieved the best results, however, its prolonged training time limits its suitability for time-sensitive applications. Conversely, the LSTM model provided the best balance between performance and efficiency, making it well-suited for stable environments. In the future, we will focus on improving model robustness through enhanced regularization and optimized architectures to strengthen generalization capabilities and ensure consistent performance across diverse datasets. Furthermore, integrating real-time data sources – such as live traffic and weather feeds – may enhance prediction accuracy and adaptability. We also plan to explore methods from explainable AI (XAI) to improve model transparency and increase user trust. Additionally, the integration of hybrid and transformer-based models may offer improved spatial-temporal learning and greater scalability.

Author contributions

Conceptualization: TD Trinh and TNT Nguyen; Methodology: TNT Huynh and HD Bui; Formal analysis and investigation: TNT Huynh and HD Bui; Writing – original draft preparation: TNT Huynh and HD Bui; Writing – review and editing: TD Trinh and TNT Nguyen; Supervision: TD Trinh.

Disclosure statement

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

References

Alghamdi, D., Basulaiman, K., & Rajgopal, J. (2022). Multi-stage deep probabilistic prediction for travel demand. *Applied Intelligence*, *52*(10), 11214–11231. https://doi.org/10.1007/s10489-021-03047-1

Ara, Z., & Hashemi, M. (2022). Predicting ride hailing service demand using autoencoder and convolutional neural network. *International Journal of Software Engineering and Knowledge Engineering*, *32*(01), 109–129. https://doi.org/10.1142/S021819402250005X

Ashkrof, P., Homem de Almeida Correia, G., Cats, O., & van Arem, B. (2022). Ride acceptance behaviour of ride-sourcing drivers. *Transportation Research Part C: Emerging Technologies*, *142*, Article 103783. https://doi.org/10.1016/j.trc.2022.103783

Battifarano, M., & Qian, Z. (2019). Predicting real-time surge pricing of ride-sourcing companies. *Transportation Research Part C: Emerging Technologies*, *107*, 444–462. https://doi.org/10.1016/j.trc.2019.08.019

Castillo, J. C., Knoepfle, D., & Weyl, G. (2017). Surge pricing solves the wild goose chase. In *Proceedings of the 2017 ACM Conference on Economics and Computation* (*EC '17*). (pp. 241–242). Association for Computing Machinery. https://doi.org/10.1145/3033274.3085098

Chen, L., Thakuriah, P., & Ampountolas, K. (2021). Short-term prediction of demand for ride-hailing services: A deep learning approach. *Journal of Big Data Analytics in Transportation*, *3*(2), 175–195. https://doi.org/10.1007/s42421-021-00041-4

Chen, Y., Cao, J., Feng, S., & Tan, Y. (2015). An ensemble learning based approach for building airfare forecast service. In *2015 IEEE International Conference on Big Data* (*Big Data*) (pp. 964–969). IEEE. https://doi.org/10.1109/BigData.2015.7363846

Chou, K. S., Wong, K. L., Zhang, B., Aguiari, D., Im, S. K., Lam, C. T., Tse, R., Tang, S.-U., & Pau, G. (2023). Taxi demand and fare prediction with hybrid models: Enhancing efficiency and user experience in city transportation. *Applied Sciences*, *13*(18), Article 10192. https://doi.org/10.3390/app131810192

Dogo, V., Garg, K., & Zheng, Y. (2020). Ride-share analysis in Boston city: A comparison between Uber and Lyft. In *Proceedings of the 21st Annual Conference on Information Technology Education* (*SIGITE '20*). Association for Computing Machinery. https://doi.org/10.1145/3368308.3415444

Eslami, G., & Ghaderi, F. (2024). A district-centric attention mechanism approach for online ride-hailing demand forecasting. *IEEE Access*, *12*, 141190–141197. https://doi.org/10.1109/ACCESS.2024.3411512

Farris, P., Yemen, G., Weiler, V., & Ailawadi, K. L. (2014). *Uber pricing strategies and marketing communications*. SSRN.

Garg, N., & Nazerzadeh, H. (2020). Driver surge pricing. In *Proceedings of the 21st ACM Conference on Economics and Computation* (*EC '20*). Association for Computing Machinery. https://doi.org/10.1145/3391403.3399476

Guo, S., Deng, B., Chen, C., Ke, J., Wang, J., Long, S., & Xu, K. (2024). Seeking in ride-on-demand service: A reinforcement learning model with dynamic price prediction. *IEEE Internet of Things Journal*, *11*(18), 29890–29910. https://doi.org/10.1109/JIOT.2024.3407119

Huang, H. (2023). Ridesharing price prediction: Exploring the strategies of dynamic pricing. *Highlights in Business*, *Economics and Management*, *10*, 106–110. https://doi.org/10.54097/hbem.v10i.7963

Huang, H. (2023). Taxi fare prediction based on multiple machine learning models. *Applied and Computational Engineering*, *16*, 7–12. https://doi.org/10.54254/2755-2721/16/20230849

Ihianle, I. K., Nwajana, A. O., Ebenuwa, S. H., Otuka, R. I., Owa, K., & Orisatoki, M. O. (2020). A deep learning approach for human activities recognition from multimodal sensing devices. *IEEE Access*, *8*, 179028–179038. https://doi.org/10.1109/ACCESS.2020.3027979

Jashwanth, K., Reddy, K. L. S. P., Snehitha, M. S., Sampath, N., & Nair, P. C. (2024). *Analyzing urban transportation services using rideshare data insights*. In*2024 IEEE 9th International Conference for Convergence in Technology* (*I2CT*). IEEE. https://doi.org/10.1109/I2CT61223.2024.10543505

Kankanamge, K. D., Witharanage, Y. R., Withanage, C. S., Hansini, M., Lakmal, D., & Thayasivam, U. (2019, 3–5 July 2019). Taxi trip travel time prediction with isolated XGBoost regression. In *2019 Moratuwa Engineering Research Conference* (*MERCon*). IEEE. https://doi.org/10.1109/MERCon.2019.8818915

Karamanis, R., Anastasiadis, E., Angeloudis, P., & Stettler, M. (2021). Assignment and pricing of shared rides in ride-sourcing using combinatorial double auctions. *IEEE Transactions on Intelligent Transportation Systems*, *22*(9), 5648–5659. https://doi.org/10.1109/TITS.2020.2988356

Kumar, M., Gupta, D. K., & Singh, S. (2021). Extreme event forecasting using machine learning models. In G. S. Hura, A. K. Singh, & L. Siong Hoe (Eds.), *Lecture notes in electrical engineering: Vol. 668. Advances in Communication and computational technology. ICACCT 2019*. Springer, Singapore. https://doi.org/10.1007/978-981-15-5341-7_115

Li, S., Yang, H., Cheng, R., & Ge, H. (2024). Hybrid deep learning models for short-term demand forecasting of online car-hailing considering multiple factors. *Transportation Letters*, *16*(3), 218–233. https://doi.org/10.1080/19427867.2023.2175420

Li, Z., Xu, H., Gao, X., Wang, Z., & Xu, W. (2024). Fusion attention mechanism bidirectional LSTM for short-term traffic flow prediction. *Journal of Intelligent Transportation Systems*, *28*(4), 511–524. https://doi.org/10.1080/15472450.2022.2142049

Liu, S., Jiang, H., & Chen, Z. (2021). Quantifying the impact of weather on ride-hailing ridership: Evidence from Haikou, China. *Travel Behaviour and Society*, *24*, 257–269. https://doi.org/10.1016/j.tbs.2021.04.002

Pham, H. V., Lam, H. P., Duy, L. N., Pham, T. B., & Trinh, T. D. (2024). An improved convolutional recurrent neural network for stock price forecasting. *IAES International Journal of Artificial Intelligence* (*IJ-AI*), *13*(3), 3381–3394. https://doi.org/10.11591/ijai.v13.i3.pp3381-3394

Poongodi, M., Malviya, M., Kumar, C., Hamdi, M., Vijayakumar, V., Nebhen, J., & Alyamani, H. (2022). New York City taxi trip duration prediction using MLP and XGBoost. *International Journal of System Assurance Engineering and Management*, *13*, 16–27. https://doi.org/10.1007/s13198-021-01130-x

Rangel, T., Gonzalez, J. N., Gomez, J., Romero, F., & Vassallo, J. M. (2022). Exploring ride-hailing fares: An empirical analysis of the case of Madrid. *Transportation*, *49*(2), 373–393. https://doi.org/10.1007/s11116-021-10180-w

Schwieterman, J. P. (2019). Uber economics: Evaluating the monetary and travel time trade-offs of transportation network companies and transit service in Chicago, Illinois. *Transportation Research Record*, *2673*(4), 295–304. https://doi.org/10.1177/0361198119839344

Silveira-Santos, T., González, A. B. R., Rangel, T., Pozo, R. F., Vassallo, J. M., & Díaz, J. J. V. (2024). Were ride-hailing fares affected by the COVID-19 pandemic? Empirical analyses in Atlanta and Boston. *Transportation*, *51*(3), 791–822. https://doi.org/10.1007/s11116-022-10349-x

Silveira-Santos, T., Papanikolaou, A., Rangel, T., & Manuel Vassallo, J. (2023). Understanding and predicting ride-hailing fares in Madrid: A combination of supervised and unsupervised techniques. *Applied Sciences*, *13*(8), Article 5147. https://doi.org/10.3390/app13085147

Sindhu, P., Gupta, D., Meghana, S., & A. K, M. (2022). Modeling Uber data for predicting features responsible for price fluctuations. In *2022 IEEE Delhi Section Conference* (*DELCON*). IEEE. https://doi.org/10.1109/DELCON54057.2022.9752864

Sriwongphanawes, K., & Fukuda, D. (2024). How do fares affect the utilization of ride-hailing services: Evidence from Uber Japan's experiments. *Asian Transport Studies*, *10*, Article 100121. https://doi.org/10.1016/j.eastsj.2023.100121

Vuong, P. H., Dat, T. T., Mai, T. K., Uyen, P. H., & Bao, P. T. (2022). Stock-price forecasting based on XGBoost and LSTM. *Computer Systems Science and Engineering*, *40*(1), 237–246. https://doi.org/10.32604/csse.2022.017685

Vuong, P. H., Phu, L. H., Van Nguyen, T. H., Duy, L. N., Bao, P. T., & Trinh, T. D. (2025). A comparative study of deep learning approaches for stock price prediction. *Digital Finance*. https://doi.org/10.1007/s42521-025-00149-0

Yousaf, K., & Nawaz, T. (2022). A deep learning-based approach for inappropriate content detection and classification of YouTube videos. *IEEE Access*, *10*, 16283–16298. https://doi.org/10.1109/ACCESS.2022.3147519

Zhang, W., Kumar, D., & Ukkusuri, S. V. (2017). Exploring the dynamics of surge pricing in mobility-on-demand taxi services. In *2017 IEEE International Conference on Big Data* (*Big Data*). https://doi.org/10.1109/BigData.2017.8258070

Zhao, J., Chen, C., Huang, H., & Xiang, C. (2023). Unifying Uber and taxi data via deep models for taxi passenger demand prediction. *Personal and Ubiquitous Computing*, *27*(3), 523–535. https://doi.org/10.1007/s00779-020-01426-y

Zhao, X., Sun, K., Gong, S., & Wu, X. (2023). RF-BiLSTM neural network incorporating attention mechanism for online ride-hailing demand forecasting. *Symmetry*, *15*(3), Article 670. https://doi.org/10.3390/sym15030670