A Machine Learning Framework for Travel Mode Classification from Sparse GPS Data

Alvin Wong [1]   Dr Xiang Yan [2]

UF Transportation Institute UNIVERSITY of FLORIDA

JUST STOP GREEN TRANSPORTATION LAB

[1] Dept. of Computer and Information Science and Engineering, [2] Dept. of Civil and Coastal Engineering, University of Florida

Background

Identifying transportation modes from GPS trajectory data can unlock important insights into efficient urban planning. The majority of GPS trajectory data is collected passively from mobile devices, often resulting in sparse GPS trajectories with low temporal resolutions, making it challenging to reliably classify transportation modes. Although dense trajectories have been studied widely, sparse trajectories are underutilized due to difficulties in analysis even though they constitute the majority of real-world data. This study introduces a machine learning framework designed to work with sparse GPS trajectories, classifying trips into non-motorized, motorized, or bus transit categories.

Methodology

For this study, we acquired two distinct datasets, an unlabeled dataset for real-world validation of our trained models and labeled dataset for model training. The unlabeled dataset was obtained from commercial data vendors, containing raw GPS data from over one million mobile devices within the Northeast Florida region. The labeled dataset was collected from various sources and was comprised of motorized, non-motorized and bus transit trajectories. Together, these sources provided a comprehensive dataset that would serve as our training and validation sets for our machine learning models.

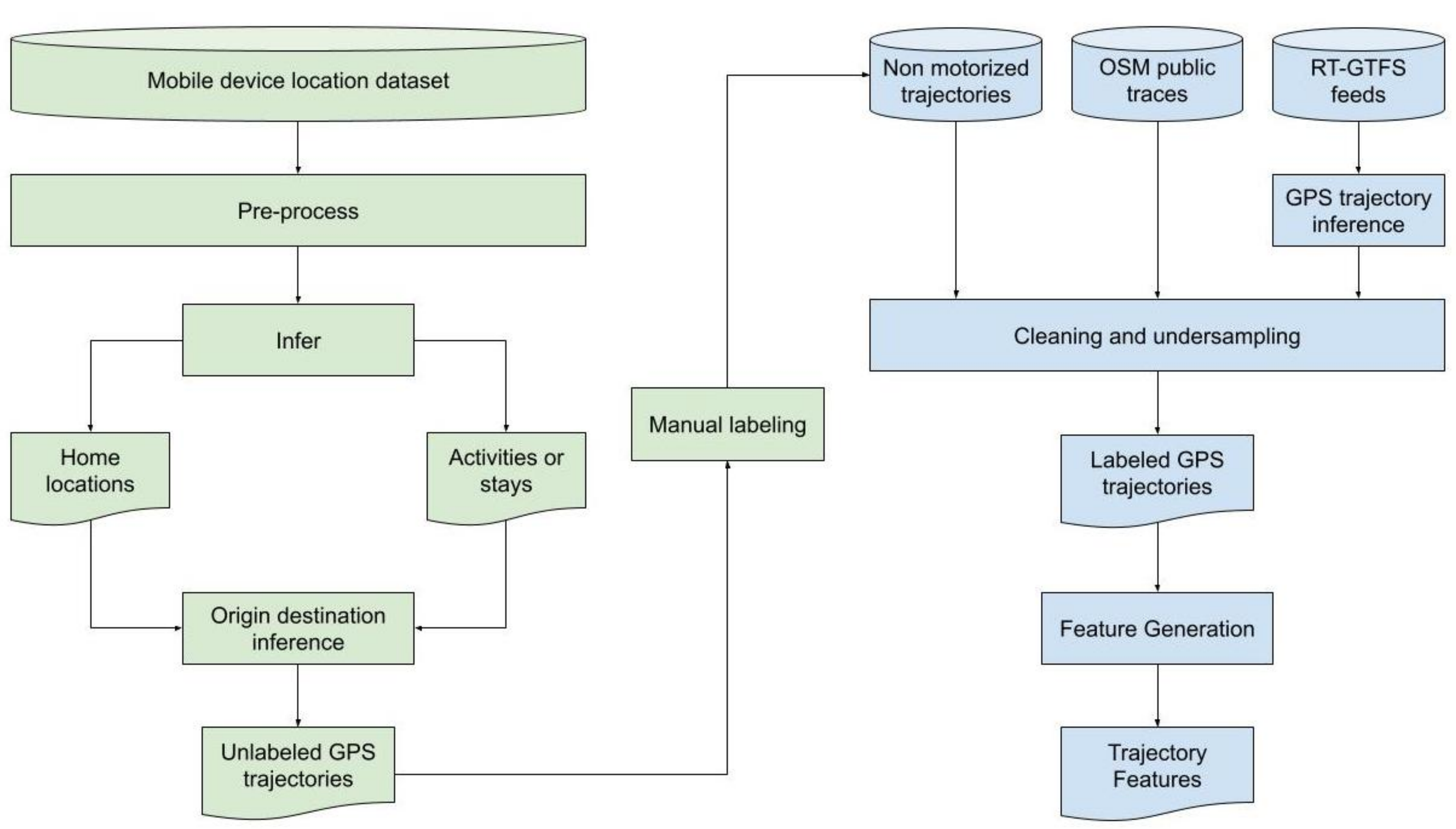

Figure 1. Data Processing Steps for GPS Trajectory Mode Inference Adapted from Lyu et al. (2025)

Once processed and cleaned, the GPS trajectories were under sampled. First, the unlabeled dataset's location recording interval (LRI) distribution was calculated. Then we iterated over each trajectory in our labeled dataset and removed GPS location points, matching the previously calculated LRI distribution. The figures below showcase the LRI distributions for the labeled trips before and after the under-sampling process.

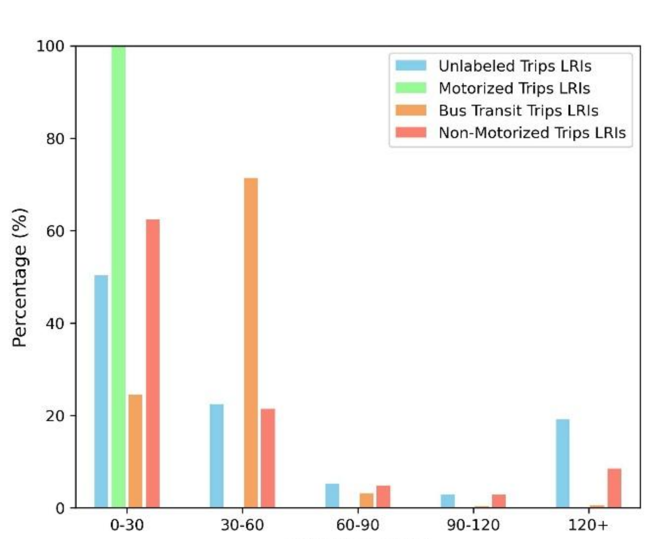

Figure 2. LRI Distribution Before Under-sampling

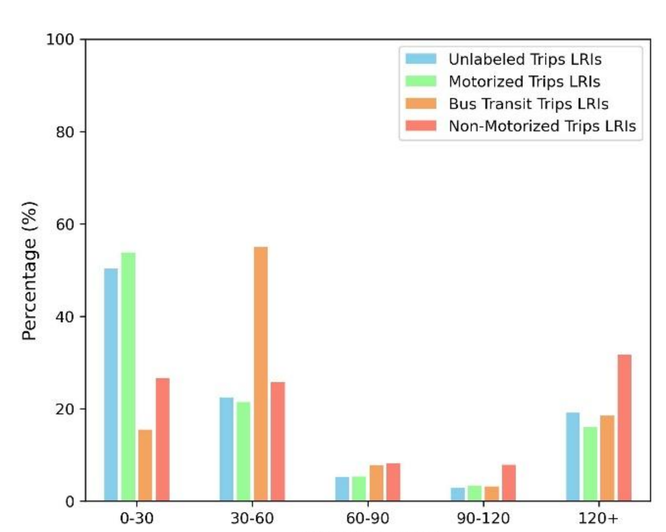

Figure 3. LRI Distribution After Under-sampling

Analysis and Results

Several spatiotemporal features were extracted such as origin-destination distance, cumulative trip distance, velocity percentiles, and distance to bus line percentiles. To provide contextual information, two additional features were added, the predicted car travel time ratio and predicted walk travel time ratio. These two features are the trip's travel time relative to the predicted car and walk travel durations based on OSRM's API.

Table 1. Features for Detecting Travel Mode Adapted from Zhang et al. (2021)

Features	Number of Variables
Origin-destination straight-line distance	1
Cumulative trip distance	1
Predicted car travel time ratio	1
Predicted walk travel time ratio	1
Average travel velocity	1
0, 5, 25, 50, 75, 95, 100 percentile travel velocity	7
0, 5, 25, 50, 75, 95, 100 percentile distances to the nearest bus lines	7

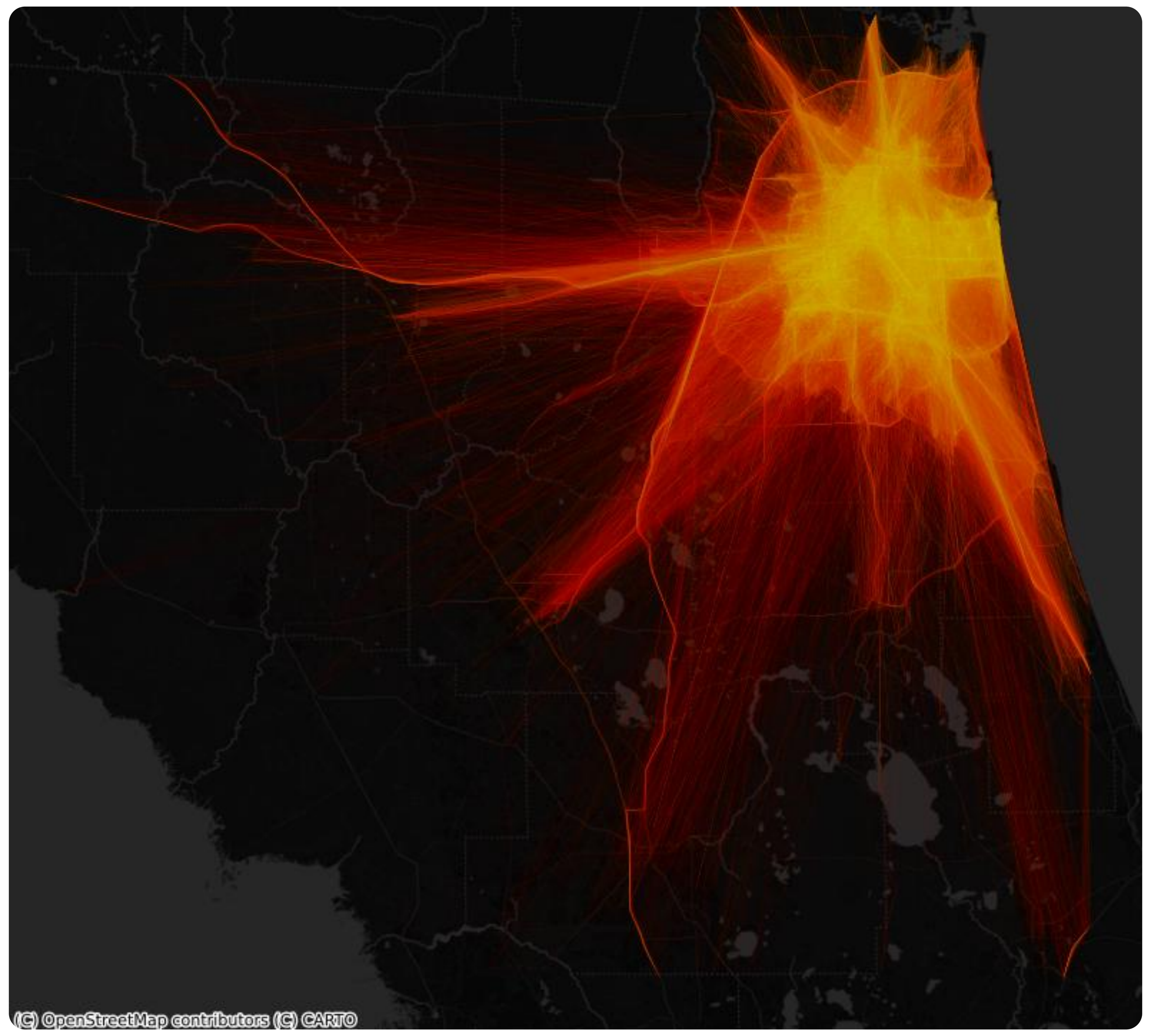

Figure 2. Heatmap of Trip Trajectories from our Mobile Device Location Dataset

Our results indicated that our final Random Forest model achieved strong performance on both the cross-validation folds and the held-out test set despite the loss of information from high LRIs. The 10-fold cross-validation process resulted in a validation accuracy of 93.5% when using 200 estimators and an unlimited maximum depth. When evaluated on the held-out test set, the model achieved a test accuracy of 94.5%.

Table 2. Classification report on the held-out test set

Class	Precision	Recall	F1-Score	Support
Non-motorized	0.9661	0.9500	0.9580	60
Motorized	0.9450	0.9450	0.9450	200
Bus transit	0.9403	0.9450	0.9426	200

Discussion

While our model performed well on the labeled datasets, validation against the real-world dataset indicated that there were issues when moving beyond our labeled data. Our model clearly overestimated the proportion of non-motorized and bus-transit trips when compared to real-world data. The difference in travel mode distribution likely arose from multiple factors. For example, the noise in our GPS data, the variability in feature generation caused by high LRIs and most importantly the small training set could be factors impacting these results. The figure below shows the predicted mode distribution compared to the 2023 Florida mode distribution.

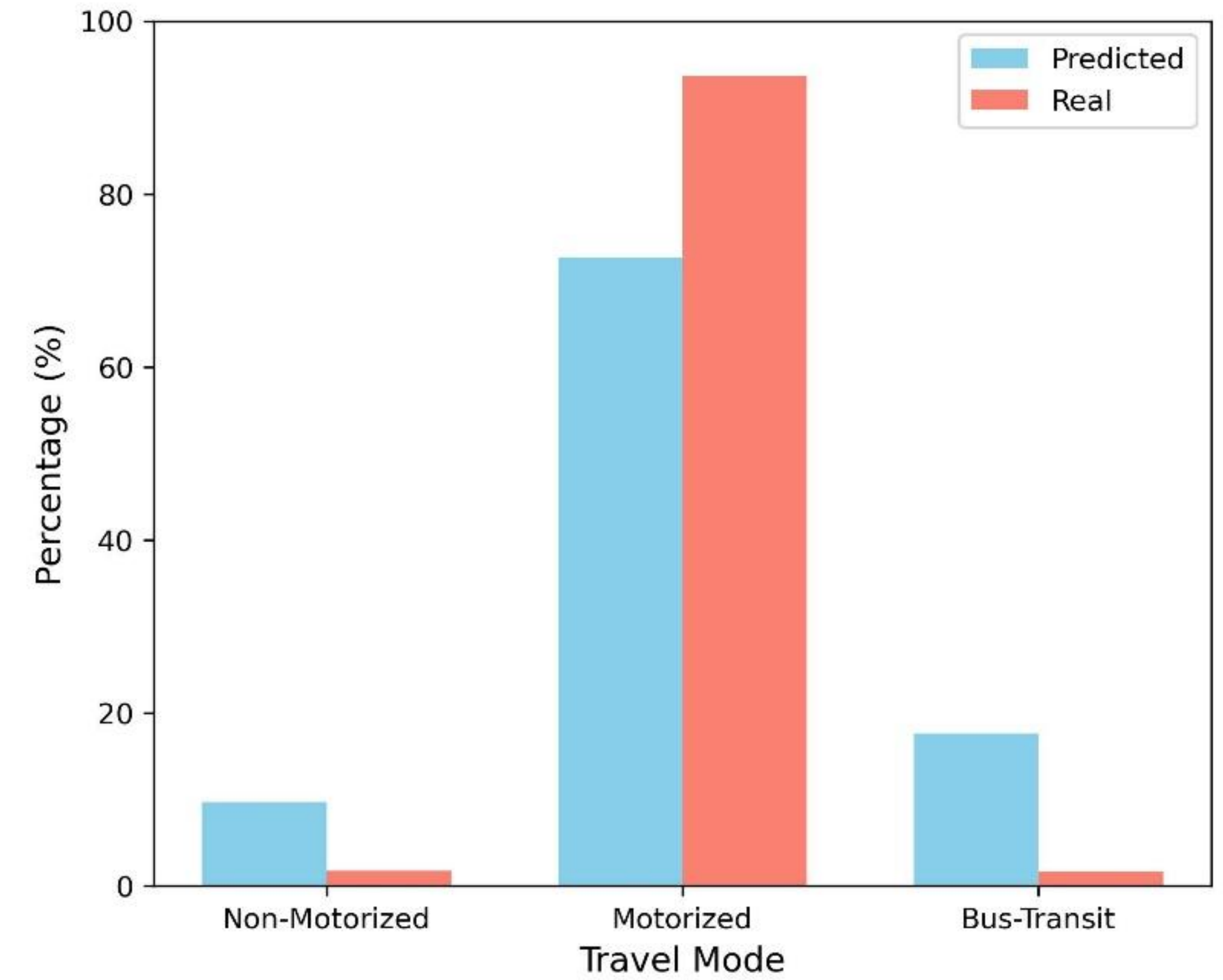

Figure 4. Predicted Travel Mode Distribution Compared to Real-World Distribution

- The machine learning framework performs well on labeled data despite the loss of information from high LRIs.
- Broadly reflects but does not match the real-world travel mode distribution, showcasing improvements are possible.

The findings from this research can have multiple practical applications in efficient urban and transportation planning. Accurately inferring modes of travel allows urban city planners to understand where residents travel and their transportation mode. This could assist in determining where transportation infrastructure is required, leading to improvements in traffic congestion and long travel times, especially in large urban areas where car travel is the main form of transportation.

References

Burkhard, O., Becker, H., Weibel, R., & Axhausen, K. W. (2020). On the requirements on spatial accuracy and sampling rate for transport mode detection in view of a shift to passive signalling data. *Transportation Research Part C: Emerging Technologies*, *114*, 99–117. https://doi.org/10.1016/j.trc.2020.01.021

Lyu, D., Liu, L., Campbell, C., Zhang, Y., & Yan, X. (2025). Potentials and limitations of large-scale, individual-level mobile location data for food acquisition analysis. https://doi.org/10.48550/arXiv.2503.18119

Yang, M., Pan, Y., Darzi, A., Ghader, S., Xiong, C., & Zhang, L. (2022). A data-driven travel mode share estimation framework based on mobile device location data. *Transportation*, 49(5), 1339–1383. https://doi.org/10.1007/s11116-021-10214-3

Zhang, L., Darzi, A., Pan, Y., Yang, M., Sun, Q., Kabiri, A., Zhao, G., Ashoori, M., Namadi, S. S., Xiong, C., Jordan, G. W., Stewart, K., & Pack, M. L. (2021). *Next Generation National Household Travel Survey National Origin-Destination Data: Passenger Origin-Destination Data Methodology Documentation*. Federal Highway Administration.

Florida Department of Transportation. (2023). Commuting trends report. https://fdotwww.blob.core.windows.net/sitefinity/docs/default-source/planning/demographic/2023-commuting-trends-report.pdf?sfvrsn=ce2b7881_1

JOURNAL OF UNDERGRADUATE RESEARCH

A Machine Learning Framework for Travel Mode Classification from Sparse GPS Data

XX

XX, *XX*

XX, XX

Abstract

Identifying transportation modes from GPS trajectory data can unlock important insights into efficient urban planning. However, real-world GPS trajectories are often sparse, collected with low temporal resolutions from mobile devices, making it challenging to reliably classify transportation modes. Although dense trajectories have been studied widely, sparse trajectories are underutilized due to difficulties in analysis despite constituting the majority of real-world data. This study introduces a machine learning framework designed specifically to work with sparse GPS trajectories, classifying trips into non-motorized, motorized, or bus transit categories. Our method down samples publicly accessible labeled datasets to simulate the low temporal resolution of real-world unlabeled GPS trajectories. We incorporate additional features that add real-world context to support our framework in handling sparse GPS trajectories. Trained on these sparse trajectories, our models are then applied to our unlabeled target dataset, and the predicted travel mode distributions are validated against real-world distributions of transportation modes. Our results indicate that our machine learning framework effectively classifies transportation modes using only sparse GPS trajectories. Our work offers a practical solution for urban planners to see where and how residents travel utilizing only sparse GPS trajectories.

Keywords: Travel mode inference, Sparse GPS trajectory data, Mobile device location data, Machine learning

Introduction

Travel mode choice is an extremely important factor in efficient urban and transportation planning, offering essential insights into improving traffic congestion, and long travel times in urban areas (Eluru et al., 2012; Dabiri and Heaslip 2018). In the past, researchers have relied largely on household surveys and interviews to gather this information. However, this approach can suffer from low response rates, high costs from labor, and inaccurate trip reporting (Stopher et al. 2007; Bricka et al. 2009). As a result, organizations have tried to find more cost-effective methods and larger datasets to capture more diverse and specific information on urban transportation.

The high prevalence and usage of GPS devices, particularly mobile devices, has emerged as an extremely promising alternative to travel mode surveys. Mobile devices can record the location and movement of travelers with higher accuracy while avoiding many of the drawbacks associated with traditional travel surveys (Chen et al., 2016; Burkhard et al., 2020; Bricka et al. 2009). Analyzing location data from mobile devices allows for the inference of travel behavior and trips that individuals take from one location to another. These trips can be represented as GPS trajectories, which this paper will define as a sequence of geographic coordinates with associated timestamps, corresponding to a travel trip of an individual. GPS trajectories enable a more detailed analysis of travel behavior, route choices, and transportation mode identification. However, many studies on identifying transportation modes from GPS trajectories often relied on dense GPS trajectory data (Burkhard et al. 2020; Yang et al., 2022). These studies utilized dedicated GPS tracking applications or devices placed on paid participants that could frequently record the device's location, offering a higher temporal resolution through a low Location Recording Interval (LRI) (Burkhard et al. 2020; Yang et al., 2022). The LRI is defined as the time interval between successive GPS points, with a low LRI providing more information in each GPS trajectory and thus higher temporal resolution and denser trajectories (Burkhard et al. 2020; Yang et al., 2022). However, real-world GPS trajectories are often sparse as many common mobile devices track location passively from sources such as location apps, background services, Wi-Fi or Call Device Records (Burkhard et al. 2020, Chen et al., 2016). Many of these collection methods typically record with a high LRI to reduce battery usage and processing or simply due to how the method of collection is performed (Burkhard et al., 2020, Yang et al., 2022, Chen et al., 2016). As a result, it is unknown if many existing inference models can identify sparse GPS trips accurately as they were designed with high temporal resolution data or require additional kinematic features such as instantaneous speed or acceleration (Burkhard et al. 2020).

Because a large majority of real-world GPS trajectories are sparse with low and variable temporal resolution, they are difficult to work with and rarely utilized in urban and transportation planning (Bolbol et al. 2012; Burkhard et al. 2020). However, developing frameworks to handle sparse GPS trajectories could unlock new insights and solutions to problems in urban planning.

Specifically, robust frameworks are required to handle the ambiguity introduced by low sampling rates, which can make distinct travel modes, such as biking compared to driving on a congested road, appear similar as data points are recorded irregularly. Additionally, real world information from travel network analysis such as roadways or transit lines can be integrated into frameworks to provide improved context when inferring travel modes. Very few studies have explored approaches for utilizing sparse GPS trajectories. Burkhard et al. (2020) tested for the required spatial and temporal resolution for travel mode inference by subsampling high temporal resolution data to have lower temporal resolution. They examined the effects of a LRI of 1s – 300s on several models, recommending that the LRI should be less than 60s for model accuracy. (Burkhard et al., 2020). Yang et al. (2022) also addressed this issue, collecting GPS trajectories with LRIs of 1, 2, 5 and 15 seconds. They applied their random forest model trained on 15 second LRI trajectories to an unlabeled mobile device location dataset and examined the travel mode distribution (Yang et al., 2022). Zhang et al. (2021) utilized the same ground truth data as Yang et al. (2022) and applied their model to a much larger unlabeled mobile device location data set. Bolbol et al. (2020) utilized an LRI of 60 seconds along with an SVM model to classify transportation modes. But despite the above studies, a large research gap still remains due to the lack of development of frameworks specifically designed for sparse GPS trajectories. Most mobile device location datasets exhibit variable and sparse temporal resolution, often exceeding LRIs of 60 seconds. This paper intends to address this research gap.

Therefore, this paper introduces a machine learning framework designed to classify travel modes from sparse GPS trajectories. Our method addresses the challenge of insufficient detail by incorporating additional contextual features and training on a simulated low temporal accuracy data set, thereby improving classification performance on sparse GPS trajectories for our three chosen target modes of non-motorized, motorized, and bus transit. A low temporal resolution is simulated by down sampling publicly accessible labeled datasets. Afterwards, multiple machine learning models are trained utilizing this data, and then applied to a target dataset of unlabeled GPS data. The results are validated against estimated travel mode distributions in the real world to assess how effectively the framework can estimate mode shares.

This study demonstrates that even with low temporal resolution, it is possible to reliably differentiate major transportation modes through a deliberate selection of features and contextual enrichment. The proposed framework can serve as a practical alternative for transportation organizations and urban planners who wish to gain insights into travel behavior from the large amount of sparse GPS trajectory data with low temporal accuracy. By focusing on common, passively collected GPS trajectory data, our approach aims to make large scale travel mode analysis more accessible and allow stronger data-driven urban and transportation planning.

Methodology

Data Acquisition

Two distinct datasets were utilized, an unlabeled dataset for validation of our trained models and labeled dataset for model training and validation. Both datasets contained GPS data points only within the North Florida region to ensure geographic consistency.

The unlabeled dataset is the exact dataset from Lyu et al. (2025) and was obtained from commercial data vendors, containing raw GPS data from over one million mobile devices within the North Florida region. This dataset offers spatial coverage over a large area while still ensuring geographic consistency to validate our trained models on.

The labeled dataset served as our ground truth data and was comprised of motorized trajectories from Open Street Map, bus transit trajectories from Real-Time GTFS feeds and manually labeled non-motorized trajectories from the unlabeled dataset. In total, 1000 motorized trajectories, 1000 bus transit trajectories and 300 non-motorized trips were collected. Together, these sources provided a comprehensive dataset that would serve as our training and validation sets for our machine learning models.

Data Processing

The unlabeled dataset utilized in this study is identical to the dataset used by Lyu et al. (2025). Therefore, we briefly summarize how they processed the raw mobile device location data to obtain GPS trajectories using the established methods of home and stay point analysis (Zhao et al., 2022; Zhang et al., 2021). First, Lyu et al. (2025) inferred home locations for device users by segmenting the area into 20-meter grids, then designating the home location as the grid with the highest number of nighttime (10:00 PM to 6:00 AM) GPS points. If users did not have nighttime GPS data, they inferred their home location using weekend daytime data (6:00 AM to 10:00 PM), under the assumption that individuals spend significant time at their residences during weekends. Afterwards, they identified stay points which are time ranges of minimal movement that are assumed to be activities. Lyu et al. (2025) then utilized the Trackintel package's time-space heuristic with a threshold of a 100-meter radius and durations of 5 to 720 minutes to detect stay points. Finally, the origins and destination points were identified through Trackintel's backward searching method to form GPS trajectories representing distinct trips.

The labeled dataset was created using a combination of Open Street Map public user traces, Real-Time GTFS feeds and non-motorized trips manually labeled from the processed unlabeled dataset. The Open Street Map user traces were already formatted as distinct GPS trajectories. The Real-Time GTFS data was processed into GPS trajectories by grouping GPS points by their vehicle ID and trip ID, dividing the data into individual bus transit trips. The non-motorized trips were manually labeled from our processed mobile device location dataset. The trips were carefully examined by using speed, origin-destination distance and contextual features. Additionally, the non-motorized trips were verified by projecting the GPS trajectories onto a map of the area. The complete data processing pipeline is shown below in Figure 1.

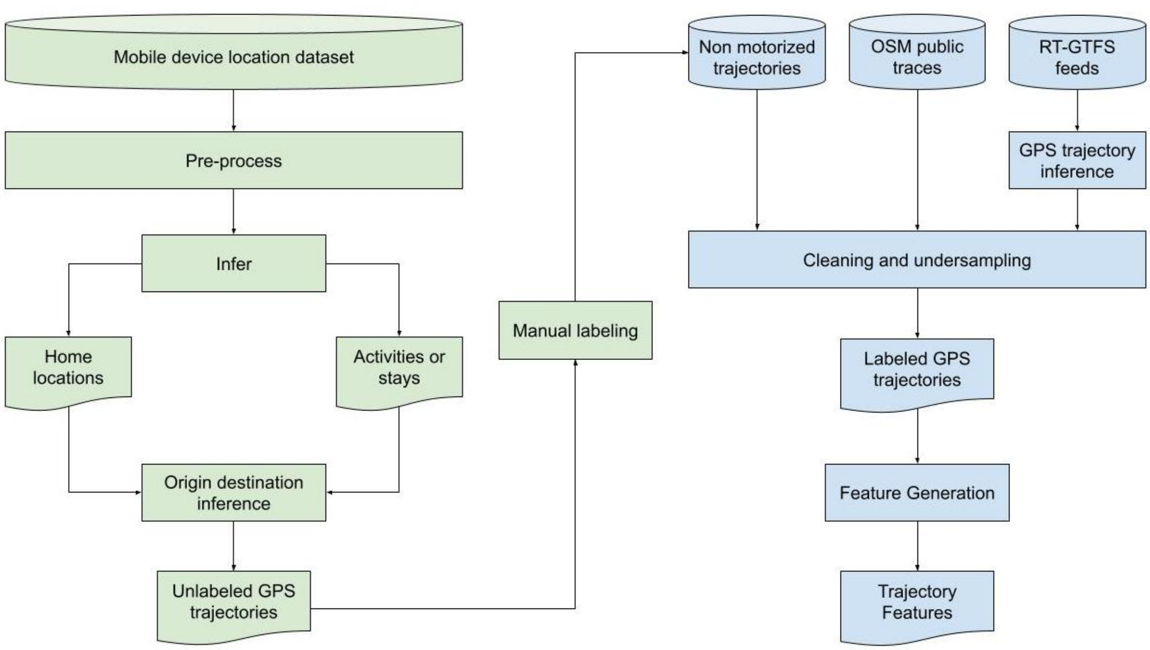

Figure 1. Data Processing Steps for GPS Trajectory Inference Adapted from Lyu et al. (2025)

Once all processed and cleaned GPS trajectories were obtained for both the labeled dataset and the unlabeled dataset, they were prepared for the under-sampling phase. First, the unlabeled dataset's LRI distribution was calculated. Then for each trajectory in our labeled dataset, multiple LRI values were randomly sampled from the unlabeled trip LRI distribution. Each trajectory was iterated over, and GPS location points were selectively removed. Afterwards, each GPS location point in the new under-sampled trajectory matched its corresponding sampled LRI value. However, because the labeled dataset's LRI values are variable, it is impossible to create an exact match with the unlabeled dataset's LRI distribution. However, the labeled dataset was able to achieve a close approximation in the distribution of LRIs. The figures below showcase the LRI distributions for the labeled trips before and after the under-sampling process.

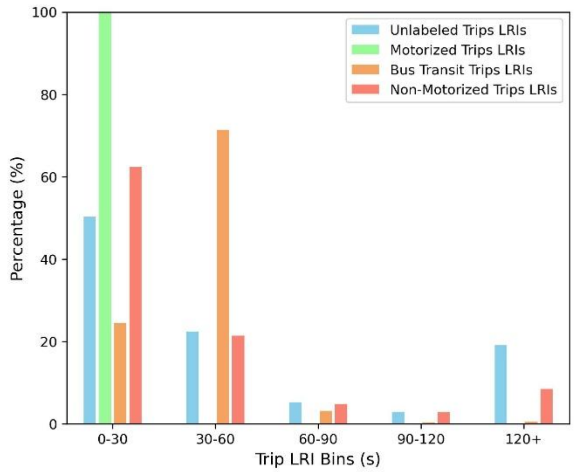

Figure 2. LRI Distribution Before Under-sampling

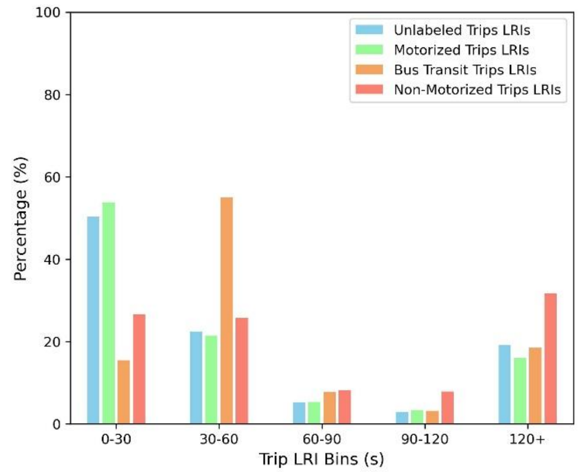

Figure 3. LRI Distribution After Under-sampling

After under-sampling, the GPS trajectories were prepared for feature extraction. Several spatiotemporal features were extracted such as origin-destination distance, cumulative distance, velocity percentiles, and the distance to bus line percentiles. These features were based on Zhang et al.'s (2021) travel inference framework. To provide contextual information, two additional features were added, the predicted car travel time ratio and predicted walk travel time ratio. These two features are the trip's travel time relative to the predicted car and walk travel durations obtained from OSRM's API. An overview of all the features utilized in training our models is summarized in Table 1.

Table 1. Features for Detecting Travel Mode Adapted from Zhang et al. (2021)

Features	Number of Variables
Origin-destination straight-line distance	1
Cumulative trip distance	1
Predicted car travel time ratio	1
Predicted walk travel time ratio	1
Average travel velocity	1
0, 5, 25, 50, 75, 95, 100 percentile travel velocity	7
0, 5, 25, 50, 75, 95, 100 percentile distances to the nearest bus lines	7

Model Training and Validation

The Random Forest was the chosen model as it outperformed other traditional machine learning models in transportation mode inference (Burkhard et al., 2020; Yang et al., 2022). The original training data had a total of 2300 labeled GPS trajectories. The Synthetic Minority Oversampling Technique (SMOTE) was applied to address the class imbalance and created synthetic data of the non-motorized minority class. This assisted in ensuring there was a balanced class distribution during the model's training. To fine-tune our model, a grid search was utilized, iterating over the "number of estimators" and "max depth" hyperparameters. To improve the model's generalizability, the 10-fold cross-validation strategy was used. An 80% training split was implemented, afterwards, each fold was trained on 90% of the training data and validated on the remaining 10%. The cross-validation accuracy served as the only metric for selecting the best hyperparameters. Our final model was trained on the full training set using the previously chosen hyperparameters and then evaluated on the held-out test set, which accounted for 20% of the labeled data. Finally, the final model was applied to the unlabeled dataset and the distribution of predicted transportation modes was compared to real-world transportation mode distributions.

Results

Our results indicated that our final Random Forest model achieved strong performance on both the cross-validation folds and the held-out test set despite the loss of information from high LRIs. The 10-fold cross-validation process had an accuracy of 93.5% when using 200 estimators and an unlimited maximum depth. When evaluated on the held-out test set, the model achieved a test accuracy of 94.5%. The table below presents the precision, recall, and F1 scores for each travel mode.

Table 2. Classification report on the held-out test set

Class	Precision	Recall	F1-Score	Support
Non-motorized	0.9661	0.9500	0.9580	60
Motorized	0.9450	0.9450	0.9450	200
Bus transit	0.9403	0.9450	0.9426	200

All three classes show strong performance, with all F1 scores remaining above 94%. We also achieved a macro-averaged F1 score of 94.9%, indicating that the model generalizes well across each travel mode. These results show that strong performance is still possible despite reduced information due to high and variable LRIs. Finally, our predicted mode distribution broadly reflects the real-world mode distribution but is not exact. The figure below shows the predicted mode distribution compared to the 2023 Florida mode distribution collected from Florida Department of Transportation (2023).

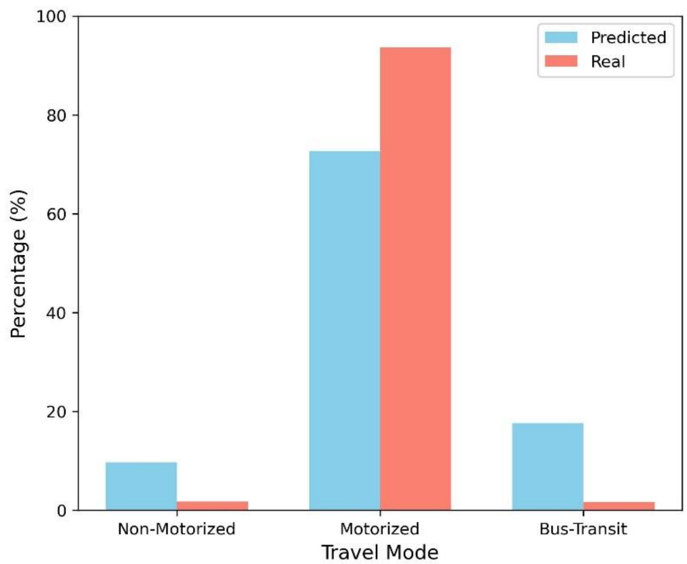

Figure 4. Predicted Travel Mode Distribution Compared to Real-World Distribution

Our model clearly overestimated the proportion of non-motorized and bus-transit trips when compared to real-world data. The difference in travel mode distribution likely arose from multiple factors. For example, the noise in our GPS data, the variability in feature generation caused by high LRIs and most importantly the small training set could be factors impacting these results. Despite this, our model captured the overall trends of real-world data, indicating that it properly utilizes relevant spatiotemporal features.

Discussion

Our final Random Forest model demonstrated strong performance with a high 10-fold cross validation and test accuracy. This indicates that even with the loss in information from high LRIs, a proper machine learning framework can discern between different travel modes. Our accuracy is comparative to previous papers that utilized dense GPS trajectory data. For example, Zheng et al. (2023) achieved a 97.1% cross-validation accuracy when utilizing 15 second LRI trajectories while Burkhard et al. (2020) achieved an ~82% test accuracy using 5 second LRI trajectories. However, the validation against the real-world dataset indicates that there are issues when moving beyond our labeled data.

One large limitation in our machine learning framework is the small size and lack of diversity of the labeled dataset. With a training dataset of only 2300 GPS trajectories across the region of North Florida, its representation of real-world travel behavior is not comprehensive. The travel behavior of the real-world data is likely to be more diverse than our training data, leading to overfitting and the discrepancies seen in the figure above. Utilizing SMOTE could have also exacerbated this bias. While SMOTE is effective in balancing the class distribution, the synthetic data may worsen the overfitting on our model. This could have also affected our model’s ability to generalize well on the unseen real-world data. To address this limitation, increasing the size and diversity of the labeled dataset is recommended to ensure the model has a more comprehensive representation of real-world data when being trained. Utilizing several data sources and more data would provide the model with a stronger ability to capture the diversity of real-world travel behavior. Additionally, incorporating additional contextual features could be beneficial, as it would reduce the model’s dependency on pure kinematic features. Including features such as the percentiles of distance to roads or walking paths could provide needed context to the model.

For future work, alternative models and technologies are recommended. Models that handle sequential data and dependencies between data, such as RNNs or Transformers, are theoretically well designed for handling the sequential nature of GPS trajectories. These models are inherently more complex than the Random Forest model and could possibly capture complicated patterns in the data that engineered features cannot. Finally, extending the geographic region of the data could provide the model with a global understanding of travel behaviors. Many research papers focus on a specific geographical region, but this limits the model's applicability, as it is unknown if it will generalize well on other geographic regions (Burkhard et al., 2020; Yang et al., 2022).

The findings from this research can have multiple practical applications in efficient urban and transportation planning. For example, accurately inferring modes of travel could allow urban city planners to understand where public transit is required and then improve public transit infrastructure in those regions. In addition, understanding where residents travel from and to as well as the form of transportation they utilize could improve traffic congestion and reduce long travel times, especially in large urban areas where car travel is the main form of transportation (Eluru et al., 2012; Dabiri and Heaslip 2018). Furthermore, many transportation agencies still rely on household surveys and interviews. Utilizing mobile device location datasets is one way to gather the same data while reducing the amount of labor required. This research is one step in the direction of utilizing mobile device location data to gain a more comprehensive understanding of travel behavior. Overall, the findings gained from this research can assist urban planners in optimizing transportation using data-driven decisions. By utilizing a robust machine learning framework instead of travel surveys and interviews, organizations can capture a more complete understanding of real-world travel patterns and plan more efficient urban transportation systems.

References

Bolbol, A., Cheng, T., Tsapakis, I., & Haworth, J. (2012). Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification. *Computers, Environment and Urban Systems*, *36*(6), 526–537. https://doi.org/10.1016/j.compenvurbsys.2012.06.001

Bricka, S., Zmud, J., Wolf, J., & Freedman, J. (2009). Household travel surveys with GPS. *Transportation Research Record: Journal of the Transportation Research Board*, *2105*(1), 51–56. https://doi.org/10.3141/2105-07

Burkhard, O., Becker, H., Weibel, R., & Axhausen, K. W. (2020). On the requirements on spatial accuracy and sampling rate for transport mode detection in view of a shift to passive signalling data. *Transportation Research Part C: Emerging Technologies*, *114*, 99–117. https://doi.org/10.1016/j.trc.2020.01.021

Chen, C., Ma, J., Susilo, Y., Liu, Y., & Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C: Emerging Technologies*, *68*, 285–299. https://doi.org/10.1016/j.trc.2016.04.005

Dabiri, S., & Heaslip, K. (2018). Inferring transportation modes from GPS trajectories using a convolutional neural network. *Transportation Research Part C: Emerging Technologies*, *86*, 360–371. https://doi.org/10.1016/j.trc.2017.11.021

Eluru, N., Chakour, V., & El-Geneidy, A. M. (2012). Travel mode choice and transit route choice behavior in Montreal: Insights from McGill University members commute patterns. *Public Transport*, *4*(2), 129–149. https://doi.org/10.1007/s12469-012-0056-2

Florida Department of Transportation. (2023). *Commuting trends report*. Author. https://fdotwww.blob.core.windows.net/sitefinity/docs/default-source/planning/demographic/2023-commuting-trends-report.pdf?sfvrsn=ce2b7881_1

Lyu, D., Liu, L., Campbell, C., Zhang, Y., & Yan, X. (2025). Potentials and limitations of large-scale, individual-level mobile location data for food acquisition analysis. https://doi.org/10.48550/arXiv.2503.18119

Stopher, P., & Greaves, S. (2007). Household travel surveys: Where are we going? *Transportation Research Part A: Policy and Practice*, *41*(5), 367–381. https://doi.org/10.1016/j.tra.2006.09.005.

Yang, M., Pan, Y., Darzi, A., Ghader, S., Xiong, C., & Zhang, L. (2022). A data-driven travel mode share estimation framework based on mobile device location data. *Transportation*, 49(5), 1339–1383. https://doi.org/10.1007/s11116-021-10214-3

Zhang, L., Darzi, A., Pan, Y., Yang, M., Sun, Q., Kabiri, A., Zhao, G., Ashoori, M., Namadi, S. S., Xiong, C., Jordan, G. W., Stewart, K., & Pack, M. L. (2021). *Next Generation National Household Travel Survey National Origin-Destination Data: Passenger Origin-Destination Data Methodology Documentation*. Federal Highway Administration.

Zhao, X., Xu, Y., Lovreglio, R., Kuligowski, E., Nilsson, D., Cova, T. J., Wu, A., & Yan, X. (2022). Estimating wildfire evacuation decision and departure timing using large-scale GPS Data. *Transportation Research Part D: Transport and Environment*, *107*, 103277. https://doi.org/10.1016/j.trd.2022.103277