

Ten quick tips for improving estimated time of arrival predictions using machine learning in logistics and transportation systems

Aasim Ayaz Wani

Department of Engineering, Cornell University, Ithaca, New York, United States

ABSTRACT

Accurate estimated time of arrival (ETA) predictions are critical for modern logistics, influencing delivery reliability, operational efficiency, and customer satisfaction across industries such as e-commerce, freight logistics, and ridesharing. This article presents ten essential strategies for improving ETA accuracy, integrating advanced machine learning techniques, real-time and historical data fusion, and traffic behavior modeling. By analyzing real-world implementations from companies like Uber, DoorDash, and Waze, we provide actionable insights for researchers and industry professionals. Our recommendations address data variability, accuracy-latency tradeoffs, and emerging challenges in dynamic transportation networks, offering a roadmap for optimizing ETA systems.

Subjects Algorithms and Analysis of Algorithms, Artificial Intelligence, Data Mining and Machine Learning, Data Science, Neural Networks

Keywords ETA prediction models, Machine learning for logistics, Deep learning for ETA accuracy, Intelligent route planning, Big data in transportation analytics, Real-time traffic forecasting, Logistics and supply chain AI, Spatiotemporal data modeling

INTRODUCTION

Accurate estimated time of arrival (ETA) predictions are central to modern logistics, mobility services, and on-demand commerce. From dock scheduling and driver dispatch to inventory staging and customer notifications, ETA forecasts shape critical operational decisions across last-mile delivery, freight networks, and ride-hailing platforms. Their economic impact is far from trivial—surveys show that nearly 70% of consumers avoid repeat purchases after delayed deliveries, and detention fees at ports can exceed \$100 per container per day once free time expires (*Amini et al., 2023; Evmidis et al., 2024*). For high-volume shippers, even a 1% systematic ETA error can translate into six-figure annual losses (*He et al., 2014*). In today's competitive landscape, precise ETA forecasting is not just a convenience—it is a strategic differentiator (*Qi & Shen, 2019; Chowdhury, Dey & Apon, 2024*).

Despite its importance, ETA modeling remains a deeply challenging task. Travel time is shaped by a complex web of interacting factors including traffic flow, road geometry, weather patterns, and driver behavior (*Jiang et al., 2023; Yuan & Li, 2021*). Urban networks introduce variability through signal timings, pedestrian interactions, and ride-hailing curb stops, while rural corridors often suffer from sparse sensing and limited rerouting options.

Submitted 21 February 2025
Accepted 11 September 2025
Published 6 October 2025

Corresponding author
Aasim Ayaz Wani,
aasimwani1@gmail.com

Academic editor
Davide Chicco

Additional Information and
Declarations can be found on
page 29

DOI 10.7717/peerj-cs.3259

© Copyright
2025 Wani

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Sudden disruptions—like construction closures, flash floods, or mass gatherings—can induce nonlinear travel time deviations that ripple across entire networks. Addressing these realities requires ETA systems that can ingest heterogeneous data sources, adapt in real time, meet strict latency constraints, and remain interpretable to both operators and regulators. However, real-world deployment is often hindered by data sparsity, privacy regulations, and the opacity of deep learning models.

In recent years, machine learning (ML) has transformed the field. Modern ETA pipelines have shifted away from heuristics toward data-driven, spatiotemporal approaches. Graph neural networks (GNNs) model road network structures with dynamic edge weights ([Li et al., 2017](#); [Pan et al., 2019](#)); Transformer architectures capture long-range temporal dependencies; and probabilistic ensembles provide uncertainty estimates essential for risk-sensitive applications ([Wen et al., 2024](#)). Hybrid approaches now combine traffic flow physics with neural residuals to better handle extreme events. Deployed systems like Uber’s *DeepETA* and DoorDash’s *DeepETA*v2 have reported up to 20% gains in long-tail accuracy while maintaining millisecond-level inference times. Nonetheless, several challenges persist, including computational overhead, bias toward over-represented corridors, and limited interpretability under high-frequency updates. While the literature on ETA modeling continues to grow rapidly, it remains fragmented across domains like transportation science, artificial intelligence, and supply-chain operations. Academic and industrial research often speak in parallel but disconnected terms, making it difficult for practitioners to consolidate best practices or identify consistent design principles. This article aims to bridge that gap.

To guide this synthesis, we adopt a scoping review methodology aimed at identifying and thematically organizing core strategies in ML-based ETA prediction. Rather than following a formal systematic review protocol like Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), we conducted structured searches across Google Scholar, IEEE Xplore, arXiv, and industry reports from logistics platforms such as Uber, Waze, and DoorDash. We used targeted search terms including “ETA prediction,” “machine learning traffic forecasting,” focusing on peer-reviewed and industry-validated works published between 2014 and 2024. Studies were included if they addressed empirical or applied aspects of ETA modeling; non-English, redundant, or methodologically vague works were excluded. Through thematic coding of various sources, we identified recurring trends, technical challenges, and best practices, which are consolidated into the ten tips that follow.

We present ten quick tips grounded in peer-reviewed research, industry case studies, and open-source systems. These tips synthesize effective strategies across data acquisition, model design, real-time adaptation, and fairness auditing. Our goal is not only to summarize what works, but also to highlight where current approaches fall short—whether due to data sparsity, concept drift, interpretability barriers, or equity risks. By translating lessons from research into actionable guidance, we aim to support developers, researchers, and logistics professionals in designing ETA systems that are accurate, scalable, and resilient to real-world complexity. The remainder of this article is organized around ten

core strategies, each focusing on a key component of ETA prediction. Together, they offer a practical roadmap for building the next generation of intelligent transportation systems.

TIP 1: THE ROLE OF ACCURATE BASE MAPS IN ETA PREDICTIONS

A base map is a digital representation of road networks and transportation infrastructure, and it serves as the foundation for ETA prediction systems ([Abdi & Amrit, 2021](#)). It encodes road geometry, connectivity, and metadata such as speed limits, tolls, and traffic regulations, thereby providing the static framework upon which dynamic data—such as live traffic and weather conditions—can be overlaid to improve route calculations and travel time predictions. Without an accurate base map, even the most advanced ETA models risk producing unreliable results. A highly accurate base map ensures that all computations within the ETA system, including routing, traffic analysis, and travel time estimation, are grounded in precise spatial data ([Wang, Fu & Ye, 2018](#)). However, modern ETA systems must go beyond static maps by integrating real-time updates, which introduces challenges related to data volatility, latency, and scalability (See [Table 1](#) for more details about base map comparisons.).

- **Accurate road network representation** is critical for improving ETA precision. High-quality base maps incorporate road geometry details, such as lane widths, curvature, and elevation, while metadata—including speed limits and intersection rules—aligns predictions with real-world conditions ([Chiang, Leyk & Knoblock, 2014](#)). Studies indicate that refining base maps with high-resolution data can reduce ETA prediction errors by up to 20%, underscoring their significant impact on model performance.
- **Topological details** further enhance base maps by capturing road features such as intersections, overpasses, and roundabouts. These elements play a crucial role in estimating travel times because congestion-prone intersections and high-occupancy vehicle (HOV) lanes significantly influence vehicle speeds. Platforms like *Google Maps* use detailed topological metadata to dynamically adjust ETAs based on lane-specific configurations and known traffic bottlenecks ([De, 2022](#)).
- **Real-time updates and continuous accuracy maintenance** are essential for keeping base maps relevant. Roads frequently undergo construction, lane closures, or realignments. Thus, without regular updates, ETA models may generate misleading predictions. Platforms such as Waze and Google Maps mitigate this issue by incorporating sensor data, satellite imagery, and crowdsourced reports ([Sasson, 2023](#)). For instance, Waze's user-driven reporting system reduces outdated map errors by leveraging live feedback from millions of drivers.
- **Integrating real-time data with static base maps:** Static base maps alone are insufficient for modern ETA systems, as real-time factors such as traffic congestion, accidents, and weather conditions introduce unpredictable variations in travel times. Therefore, effective ETA models must integrate dynamic inputs while maintaining computational efficiency ([Yang et al., 2022](#); [Pan et al., 2019](#)). Volatility in real-time data streams arises

from sudden disruptions, such as multi-vehicle collisions or extreme weather events. These anomalies create ripple effects that impact surrounding roads, necessitating dynamic adjustments by ETA systems. For example, *Uber's DeepETA* integrates real-time traffic feeds to correct prediction deviations during periods of high traffic variability (Pan et al., 2019). Latency in real-time data processing is another challenge. Delays in updating congestion levels can result in outdated ETAs. For instance, if traffic conditions are updated only every few minutes, a rapid build-up of congestion may not be captured in time. To address this, edge computing frameworks process traffic updates closer to users, which reduces data transmission delays and improves ETA reliability (Kumar et al., 2021).

- **Scalability and localization** are critical considerations for ETA systems deployed across diverse regions. Global ride-hailing and logistics platforms must tailor their base maps to regional driving behaviors and infrastructure variations. For example, urban environments may exhibit aggressive lane-switching behavior and congestion-prone intersections, while rural areas may lack updated speed limit data due to infrequent mapping updates (Shippeo, 2024). Google Maps and Uber leverage regional embeddings to adapt predictions based on localized driving patterns, while edge-cloud architectures enable the lightweight processing of real-time updates at a regional level (Chowdhury, Dey & Apon, 2024).
- **Advanced modeling techniques** play a key role in merging real-time data with static base maps. Graph based models road networks as dynamic graphs, where nodes represent intersections and edges encode attributes like speed limits and congestion levels. Consequently, these networks continuously adjust travel times based on live updates, ensuring more accurate predictions (Jiang et al., 2023; Battaglia et al., 2018). Additionally, multi-resolution grids optimize spatial data representation, allowing models to efficiently process both local and global routing dynamics (Guo et al., 2019).

In conclusion, reliable ETA prediction depends on base maps that integrate both static and real-time data. Without continuous updates, even advanced models risk obsolescence under rapidly changing conditions (Niu & Silva, 2020; Yang et al., 2022). Accurate and dynamic base maps enable precise routing by reflecting up-to-date road network conditions. Advances in satellite imagery further enhance their scalability and precision, maintaining their relevance in increasingly complex transportation systems (Yang et al., 2022; Rahman, Abdel-Aty & Wu, 2021).

TIP 2: IMPACT OF TRAFFIC BEHAVIOR ON ETA PREDICTIONS

Accurately estimating arrival times in dynamic traffic environments represents a core challenge in the domain of intelligent transportation systems (ITS), with widespread applications across logistics, ride-hailing, and urban mobility planning. Variability in road conditions, recurring congestion, and exogenous disruptions such as accidents or weather events render ETA prediction a nontrivial task. Traditional approaches that rely on static shortest-path algorithms are often insufficient in capturing stochastic traffic behavior. In

response, researchers have adopted advanced methods—including deep learning, probabilistic forecasting, and multi-source data fusion—to improve model precision, responsiveness, and robustness. Nevertheless, critical challenges persist, particularly in the localization of traffic behavior, scalability of prediction systems, and maintenance of real-time responsiveness in diverse urban and rural contexts. [Table 2](#) summarizes how different traffic behaviors influence ETA performance.

A nuanced understanding of traffic behavior is essential for enhancing ETA accuracy. Traffic flow is shaped by a complex interplay of systemic and stochastic factors, including congestion cycles, human driving decisions, and external influences such as roadwork and adverse weather. These elements produce significant fluctuations in travel time distributions, demanding predictive models capable of capturing non-linear, spatiotemporal dependencies. Scalable models must therefore optimize the trade-off between computational tractability and predictive granularity while accommodating localized mobility patterns across heterogeneous infrastructure. Several key factors affecting ETA prediction accuracy are detailed below, each requiring targeted modeling strategies:

- **Fluctuating traffic conditions:** The inherently nonstationary nature of traffic—affected by commuting patterns, infrastructure bottlenecks, and disruptions such as collisions or lane blockages—challenges predictive reliability. To mitigate this, real-time traffic ingestion and anomaly detection mechanisms are increasingly integrated into ETA pipelines. The fusion of deterministic models for structured flows with probabilistic methods (*e.g.*, Bayesian inference, deep temporal networks) has shown promise in capturing short-term variability ([Sasson, 2023](#); [Uber AI, Data/ML, 2022](#)).
- **Spatiotemporal dependencies:** Traffic systems exhibit intricate temporal rhythms and spatial correlations shaped by peak-hour usage, road geometry, and dynamic rerouting. Graph-based models such as GNNs offer a structured framework for learning these dependencies ([Shi et al., 2024](#)). Additionally, multi-agent reinforcement learning approaches can simulate vehicle-to-vehicle interactions and congestion feedback mechanisms, further refining real-time adaptability.
- **Latency-induced discrepancies:** Real-time data sources often introduce delays or inconsistencies due to variable update intervals. These temporal misalignments can propagate inaccuracies within the ETA estimation pipeline. Dynamic time warping and recurrent fusion models have been employed to temporally synchronize disparate data streams and reduce prediction lag ([Mondal, 2022](#)).
- **Environmental disruptions:** Road conditions are frequently influenced by exogenous factors such as extreme weather, construction, or special events. The integration of auxiliary datasets (*e.g.*, meteorological feeds, infrastructure metadata) enhances forecasting robustness ([Gupta, Gulla & Mancini, 2023](#); [Shippeo, 2024](#)). Convolutional neural networks (CNNs), applied to traffic camera streams, further aid in dynamically assessing road state to fine-tune ETAs.
- **Urban-rural contrast:** Urban environments tend to exhibit congestion induced by high-density intersections and signalized corridors, while rural areas often suffer from

Table 1 Comparison of static and dynamic base maps for ETA predictions.

Feature	Static base maps	Dynamic base maps	Frequency
Road geometry representation	Fixed networks, limited lane detail; often outdated.	Continuously updated with lane widths, elevation, new roads.	6 months–1 s
Route information	Planned routes based on historic data.	Real-time rerouting based on incidents, congestion.	3 months–1 min
Update frequency	Monthly or yearly updates.	Real-time or near real-time updates.	1 year–1 s
Accuracy impact	Susceptible to outdated road info.	Adjusts instantly for closures, detours, construction.	6 months–30 s
Use cases	Suitable for stable, rural road networks.	Critical for urban, high-change areas.	1 year–1 h
Traffic conditions	Relies on historical congestion patterns.	Integrates live traffic reports and closure alerts.	3 months–1 min
Weather influence	Based on seasonal trends only.	Uses real-time weather APIs (rain, fog, snow).	6 months–1 h
Vehicle factors	Fixed fuel assumptions, static vehicle specs.	Tracks real-time fuel, EV battery, and health.	1 year–5 min
Driver behavior	Based on historical speed averages.	Real-time capture of speed, braking, fatigue.	1 year–10 s
Delivery constraints	Fixed schedules and buffer logic.	Recalculates priorities, adjusts ETA on-the-fly.	6 months–5 min
External disruptions	Ignores unplanned events (e.g., protests).	Adjusts for emergencies, real-time disruptions.	1 year–5 min
Examples	OpenStreetMap, static GIS maps.	Google Maps Live, Waze, Uber Movement.	1 s

data sparsity due to limited sensor coverage. Federated learning allows for decentralized training across geographically distributed data silos without compromising privacy, addressing sparsity in rural settings ([Uber AI, Data/ML, 2022](#)). Data augmentation using synthetic trajectories also contributes to model generalization in low-data environments.

- **Heterogeneous spatial representations:** ETA models often struggle with inconsistent road network representations arising from discrepancies across mapping providers. Variability in granularity, edge segmentation, and topology necessitates harmonization through spatial embedding and alignment strategies. Techniques such as graph embedding and topological normalization improve interoperability and reduce integration errors ([Bast et al., 2016](#)).
- **Cross-regional adaptability:** ETA models trained in one city often exhibit degraded performance when deployed elsewhere due to domain shifts in traffic behavior, infrastructure design, and regulatory norms. Transfer learning and domain adaptation frameworks address this by fine-tuning models on region-specific data while preserving shared spatiotemporal features ([Naik, 2024](#)).
- **Integration of multimodal transportation:** With the rise of multimodal systems—including ride-hailing, public transit, and micromobility—ETA prediction must account for mode-switching behaviors and intermodal interdependencies. Accurate ETA models must incorporate dynamic transitions between transport modes, enabling seamless predictions across hybrid networks.

In conclusion, integration of deep learning, graph-based architectures, and real-time data fusion continues to advance the state-of-the-art in ETA prediction. As cities adopt more complex and interconnected transit ecosystems, future research will need to address challenges related to federated learning scalability, fine-grained spatiotemporal

embeddings, and the simulation of multimodal traffic environments. Progress in these areas will be pivotal to achieving real-time, transferable, and highly accurate ETA systems in intelligent mobility infrastructures.

TIP 3: MODELING DRIVER BEHAVIOR FOR ETA PREDICTIONS

Driver behavior is a fundamental determinant of travel time variability and introduces stochastic elements that challenge traditional ETA models. Unlike static infrastructure constraints, driver behavior dynamically adapts to traffic flow, road conditions, and cognitive fatigue, which makes real-time adaptability essential for accurate ETA predictions. Variability in driver decision-making contributes to 15–30% of ETA deviations, particularly in mixed urban-highway environments ([Abdi & Amrit, 2021](#)). Factors such as speed variability, lane-switching tendencies, braking patterns, and cognitive stress responses significantly influence travel times. To mitigate these uncertainties, modern ETA systems integrate real-time telematics, behavioral clustering, and probabilistic adaptation techniques. (See [Fig. 1](#) and [Table 3](#) for details).

- Speed and acceleration variability:** Variability in driver-specific speed regulation and acceleration patterns is a major determinant of both average travel time and its variance. Highway segments generally exhibit smoother and more predictable acceleration profiles, whereas urban contexts necessitate frequent deceleration due to traffic signals, pedestrian crossings, and congestion ([Munigety & Mathew, 2016](#)). Segmenting drivers into behavioural archetypes—such as cautious, moderate, and aggressive—has been shown to improve ETA prediction accuracy by up to 12% ([Pan et al., 2019](#)). Aggressive drivers tend to exceed speed limits, perform rapid accelerations, and initiate late braking, which can reduce mean travel times but introduces higher unpredictability, particularly in dense traffic ([Szumska & Jurecki, 2020](#)). In contrast, cautious drivers maintain extended following distances and exhibit smoother acceleration-deceleration cycles, promoting consistency at the expense of longer travel durations. Adaptive drivers dynamically adjust their behavior in response to contextual cues—such as traffic flow and road geometry—thereby optimizing efficiency while minimizing variance ([Singh & Kathuria, 2021](#)). Commercial platforms such as Uber and Lyft leverage real-time telematics and historical driving data to tailor ETA predictions to individual driver profiles. By integrating speed telemetry into routing algorithms, these systems mitigate systemic underestimation for conservative drivers and enhance overall prediction robustness.
- Lane-changing behavior and ETA variability** Frequent lane-switching significantly alters travel times, particularly in congested environments. While lane changes can provide faster routes by avoiding bottlenecks, they also introduce braking-induced delays for surrounding vehicles. Research shows that lane-switching tendencies increase ETA variance by 10–15%, especially in urban traffic networks ([Yang et al., 2022](#)). To model the impact of lane-switching more effectively, modern ETA frameworks employ graph-based trajectory modeling techniques, to capture real-time adjustments. These

Table 2 Impact of traffic behavior on ETA accuracy. This table outlines key traffic-related factors affecting ETA prediction, along with their effects and mitigation strategies.

Traffic factor	Impact on ETA accuracy	Data insights & mitigation strategies
Congestion (Peak hours)	Severe ETA underestimation due to bottlenecks and unpredictable slowdowns.	Real-time sensors, GPS, and ML congestion prediction; Reinforcement Learning (RL)-based rerouting.
Spatial variability	Moderate ETA errors on rural roads due to limited detour options.	Historical delay modeling, satellite imagery, hybrid offline/online ETA models.
Temporal variability	Large errors during weekends or non-commute hours due to traffic irregularity.	Time-series forecasting with Bayesian calibration for anomaly periods.
Public events	Major localized disruptions often missed by standard models.	Event calendars, crowdsourced data, predictive learning from past disruptions.
Long-distance freight routes	ETA variability from tolls, stopovers, and load-based delays.	Truck-specific GPS datasets and regulatory-aware route optimizations.
Highway vs. Urban traffic	ETA distortion from highway bottlenecks or urban stop delays.	Multi-modal routing with lane-level and zone-specific ETA adjustments.
Traffic spillover effects	Secondary congestion undermines rerouting effectiveness.	Flow modeling at intersections; predictive AI to avoid cascade effects.
Seasonal variability	Holiday peaks and weather shifts increase ETA uncertainty.	Seasonal trend modeling combined with real-time event feedback.

models analyze historical lane-switching patterns to adjust ETA predictions dynamically, compensating for driver-specific tendencies toward frequent or hesitant lane changes. Google Maps and Waze incorporate similar strategies, utilizing crowdsourced driver trajectory data to estimate the probability of lane-switching in high-traffic zones. This method improves ETA stability in highway merging areas, where aggressive lane changes can lead to significant travel time deviations.

- **Braking patterns as a proxy for driver stress and fatigue.** Braking intensity serves as a key proxy for traffic congestion, driver stress levels, and cognitive fatigue. Hard braking events are often indicative of stop-and-go traffic conditions, cognitive fatigue that delays reaction times, or aggressive driving behavior, which may temporarily reduce travel times but increase variability due to frequent deceleration events. The integration of braking telemetry into ETA models has enabled the real-time recalibration of predicted travel times, reducing errors by 9–11% in high-variance traffic conditions ([Jiang et al., 2023](#)). For instance, Amazon’s Zoox Robotaxi employs braking telemetry as a core feature in its autonomous driving ETA model ([O’Neill, 2022](#)). By analyzing braking intensity in real-world road tests, the system dynamically adjusts ETA predictions based on different road types and driving behaviors.

Empirical studies converge on the conclusion that driver heterogeneity materially impacts ETA precision; however, methodological approaches vary. Telematics-based clustering (k-means, Gaussian mixture) yields interpretable archetypes but may oversimplify continuous behavioural spectra. Deep representation learning (e.g., behaviour embeddings in DeepETA ([Uber AI, Data/ML, 2022](#))) captures subtler nuances yet demands extensive labelled data and raises privacy concerns. Hybrid pipelines—combining interpretable clustering for coarse adjustment with deep sequential

models for fine-grained correction—currently offer the best trade-off between transparency and predictive power.

Incorporating real-time behavioural telemetry into ETA models demonstrably reduces prediction error across diverse driver populations. Future research should (i) explore federated learning frameworks to preserve driver privacy while leveraging cross-fleet data; (ii) integrate physiological proxies (*e.g.*, eye-tracking, heart-rate variability) to capture cognitive fatigue more accurately; and (iii) develop transfer-learning schemes that generalise behavioural models across regions with differing regulatory environments.

TIP 4: REAL-TIME ADAPTABILITY IN FLEET ETA PREDICTIONS

Accurate ETA prediction in fleet-based transportation systems requires real-time adaptability to dynamic traffic conditions, continuous vehicle monitoring, and individualized modeling of driver behavior. Traditional static routing approaches—predominantly reliant on historical averages and fixed traffic assumptions—frequently underperform in the presence of real-world disruptions such as accidents, construction, or demand surges. In contrast, modern ETA frameworks integrate real-time telematics, sensor fusion, and adaptive decision-making pipelines, allowing for continuous model recalibration in response to evolving road conditions and stochastic perturbations (*Liu et al., 2024*) (See [Fig. 2](#) for a comparative illustration of static *versus* dynamic map integration.).

This real-time adaptability is crucial for mitigating inefficiencies, optimizing route allocations, and reducing delays arising from congestion, infrastructure changes, or driver fatigue (*Li, He & Wu, 2022*). Industry platforms such as Uber Freight, Lyft, and DoorDash exemplify this paradigm, employing advanced machine learning techniques, fatigue-aware analytics, and dynamic routing algorithms to enhance ETA precision at scale. These systems leverage deep learning architectures, historical trajectory embeddings, and multimodal sensor data to produce accurate, context-sensitive arrival time estimates across large and heterogeneous transportation networks. A comparative summary of major ETA modeling techniques, highlighting their respective strengths, weaknesses, and deployment constraints, is presented in [Table 4](#).

- Fatigue detection and rest break optimization:** Driver fatigue represents a significant source of ETA variability in long-haul logistics, arising from prolonged driving durations, circadian misalignment, and environmental stressors. Cognitive exhaustion impairs reaction times and increases the likelihood of unscheduled stops or deviations, thus impacting travel time reliability. Regulatory mandates, such as hours of service (HOS) laws, prescribe rest intervals to mitigate these effects; however, empirical data reveal substantial heterogeneity in compliance, influenced by factors including individual circadian cycles, cumulative workload, and weather conditions. To capture the stochastic nature of fatigue-induced slowdowns, modern fleet management systems increasingly integrate multimodal sensor streams—such as biometric telemetry, ocular tracking, and neurocognitive reaction-time assessments—to

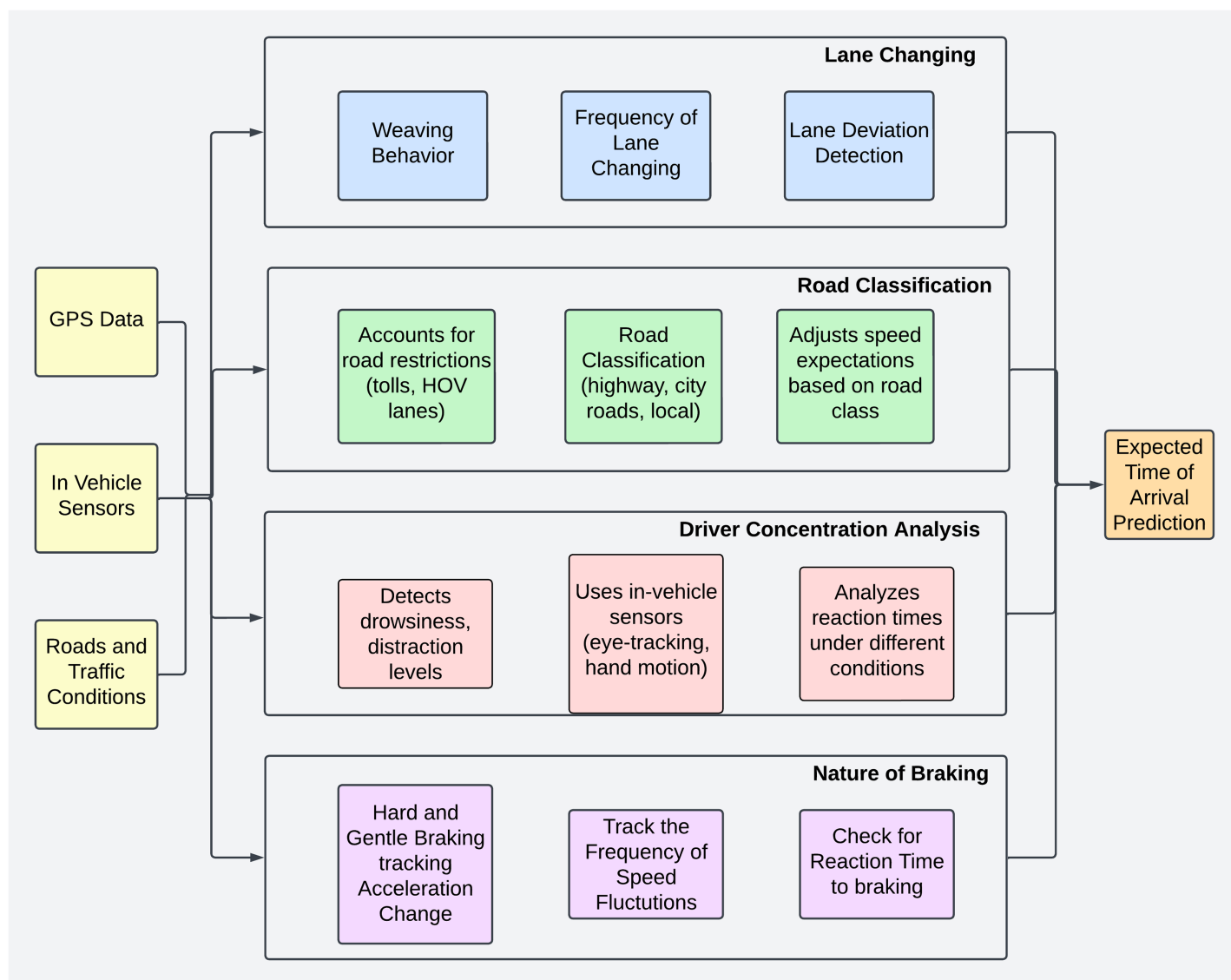


Figure 1 This flowchart illustrates the integration of driver behavior, road classification, and external conditions into ETA prediction. The system processes GPS data, in-vehicle sensor readings, and real-time road and traffic conditions to refine travel time estimates. Lane-changing behavior, including weaving, lane deviation, and frequency of lane changes, is assessed to gauge traffic fluidity. Road classification accounts for tolls, road types, and speed adjustments based on infrastructure. Driver concentration analysis relies on in-vehicle sensors to detect drowsiness, distractions, and reaction times, influencing driving efficiency. The nature of braking is monitored through acceleration changes, speed fluctuations, and reaction times, providing insights into driving patterns. These inputs collectively enhance the ETA prediction model by adapting to real-time conditions and ensuring more accurate and dynamic travel time estimates.

Full-size DOI: [10.7717/peerj-cs.3259/fig-1](https://doi.org/10.7717/peerj-cs.3259/fig-1)

infer probabilistic fatigue onset. Through real-time sensor fusion and predictive analytics, these systems support proactive interventions, including dynamic rest scheduling and route adjustment, aimed at mitigating fatigue-related delays.

Commercial implementations underscore the efficacy of this approach: Amazon's Zoox platform incorporates real-time driver monitoring to recalibrate schedules dynamically, achieving up to an 18% reduction in fatigue-related ETA deviations ([Rahman, Abdel-Aty](#)

Table 3 Impact of driver behaviors on traffic flow and ETA prediction accuracy.

Driver behavior	Traffic impact on ETA	Error (%)	Likelihood	Mitigation
Drowsy/Fatigued driving	Slower speeds and reaction time; ETA often underestimates trip duration.	15–30%	Occasional	Fatigue detection, rest mandates
Aggressive driving	Speed spikes and higher accident risk; ETA underestimates true arrival time.	10–25%	Frequent	Speed monitoring, adaptive cruise control
Distracted driving	Inconsistent motion; poor lane changes and delayed reactions reduce ETA reliability.	10–25%	Frequent	Driver monitoring, hands-free rules
Hard braking/Stops	Stop-and-go driving lowers efficiency; ETA overestimates speed consistency.	10–20%	Occasional	Driver coaching, braking assistance
Inexperienced drivers	Hesitant driving and braking disrupt flow; ETA overestimates speed.	10–20%	Occasional	Driver training, lane assist tools
Frequent lane changes	Traffic instability and unpredictable slowdowns; ETA loses precision.	5–15%	Frequent	Lane discipline programs
Strict law adherence	Predictable but slightly slow; ETA may miss delays from compliance.	5–10%	Frequent	Not applicable (positive behavior)

Wu, 2021). Likewise, Uber Freight employs machine learning–driven fatigue modelling to continuously refine ETA estimates in long-haul freight contexts, enhancing both prediction accuracy and regulatory compliance.

- Personalized routing heuristics and algorithmic deviations.** While navigation algorithms optimize fleet routing based on shortest path heuristics and real-time traffic data, empirical analyses reveal that experienced drivers frequently deviate from algorithmically recommended routes. These deviations, often driven by heuristically learned shortcuts, road familiarity, and traffic intuition, introduce variance into ETA models that fail to account for human route selection biases. Graph-based trajectory modeling techniques, provide an advanced methodological framework for embedding driver-specific routing patterns into ETA predictions. By parameterizing historical deviation behaviors, these models dynamically adjust route probability distributions and refine the fidelity of travel time estimations under high-variance conditions. Additionally, reinforcement learning-based adaptive routing frameworks iteratively refine driver-specific recommendations over extended operational durations, incrementally improving prediction precision. Empirical studies indicate that hybrid methodologies, integrating algorithmic routing intelligence with human heuristics, improve ETA prediction accuracy by 15–20.
- Variance-aware ETA predictions and probabilistic modeling.** Traditional ETA models generate point estimates that fail to capture the inherent uncertainty in driver behavior and external disruptions. To address this limitation, probabilistic modeling techniques, such as Gaussian mixture models and Bayesian inference frameworks, are employed to represent travel time distributions as continuous probability densities rather than fixed point values. Furthermore, multi-agent reinforcement learning (MARL) frameworks enable fleet-wide optimization of travel time predictions by modeling vehicle interactions as a collaborative system. Unlike conventional ETA models, which treat

drivers as independent entities, MARL-based architectures optimize individual and collective ETAs, thereby reducing congestion-induced errors by 9–14% ([Chowdhury, Dey & Apon, 2024](#)).

- **Fleet-wide coordination and multi-agent optimization:** In large-scale fleet operations, ETA inaccuracies frequently arise from inter-vehicle dependencies, congestion feedback loops, and imbalances between supply and demand. Addressing these systemic challenges requires collaborative optimization frameworks capable of real-time, distributed coordination across agents (*e.g.*, vehicles, depots, and control centers). Multi-agent systems enable vehicles to exchange and assimilate route performance metrics dynamically, enhancing global situational awareness and responsiveness. Leading platforms such as Uber Freight and DoorDash leverage federated learning architectures to aggregate driver-specific travel time data while upholding privacy constraints. These decentralized training approaches obviate the need for centralized data storage and reduce exposure to security vulnerabilities, all while maintaining model fidelity. Empirical studies have demonstrated that cooperative ETA prediction models—when integrated into fleet-wide decision-making—improve prediction accuracy and increase resilience to stochastic transport disruptions. By modelling real-time inter-vehicle interactions and incorporating telematics data streams, these systems support the transition toward adaptive, data-driven logistics operations that scale across dynamic and heterogeneous transportation networks.

In conclusion, future ETA frameworks will hinge on real-time telematics, probabilistic modeling, and reinforcement-based fleet coordination. Key research priorities include privacy-preserving driver models, fatigue-aware sensor fusion, and multi-agent optimization. As models grow more adaptive, the convergence of AI, real-time data, and uncertainty modeling will drive more resilient and accurate fleet-wide ETA predictions.

TIP 5: ACCOUNTING FOR WEATHER CONDITIONS IN ETA PREDICTIONS

Weather conditions significantly influence travel times and, consequently, ETA predictions. Factors such as rain, snow, fog, and extreme temperatures, along with their indirect effects on vehicle performance and driver behavior, disrupt normal traffic flows. This disruption necessitates robust mechanisms to model and incorporate weather effects into ETA systems. Accounting for these primary and secondary influences is critical for delivering accurate and reliable predictions. (See [Fig. 3](#) for more details about the weather API data ingestion process).

- **Rain and snow** are among the most disruptive weather conditions. Heavy rain reduces visibility, makes road surfaces slippery, and increases braking distances, leading to slower vehicle speeds and a heightened risk of congestion. Similarly, snow and ice create hazardous driving conditions, forcing drivers to adopt lower speeds and potentially causing road closures. To address these effects, modern ETA systems integrate real-time weather data from Application Programming Interface (API) services and combine it

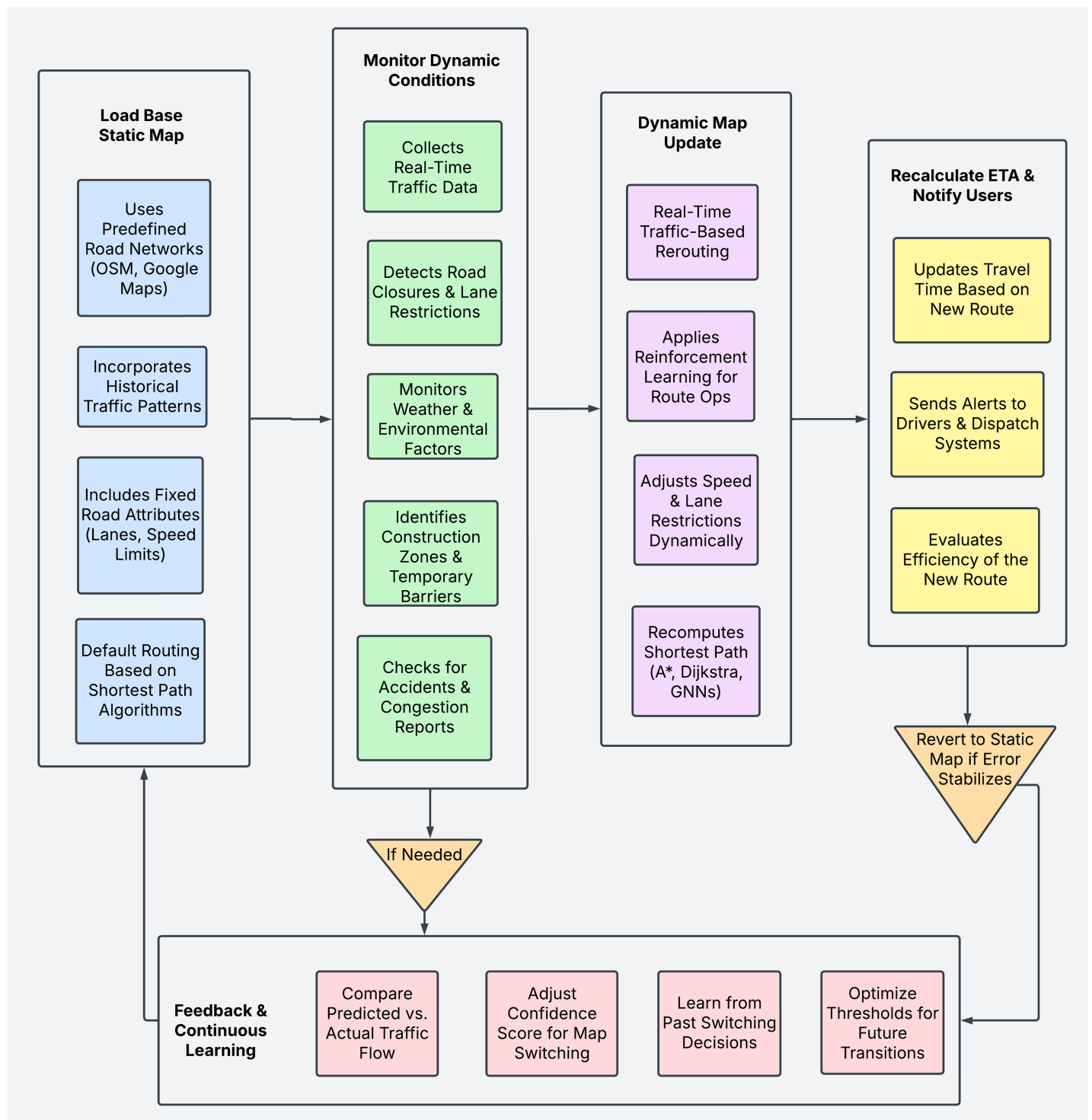


Figure 2 This flowchart outlines the decision-making process for switching between a predefined static map and a real-time dynamic map to enhance ETA accuracy. The system starts with the Static Map, using historical traffic data and shortest path algorithms. Dynamic conditions such as traffic congestion, road closures, and weather changes are continuously monitored. If disruptions exceed a predefined threshold, the system switches to a Dynamic Map, leveraging real-time rerouting, reinforcement learning, and adaptive speed adjustments. Once the new route is computed, the system recalculates ETA, notifies users, and evaluates efficiency. If traffic stabilizes, it reverts to the Static Map to optimize performance. The

Figure 2 (continued)

Feedback & Continuous Learning module refines switching decisions, adjusts confidence scores, and optimizes thresholds, ensuring future transitions are more accurate and adaptive. This approach balances computational efficiency with real-time responsiveness, improving travel time predictions and user experience.

Full-size  DOI: 10.7717/peerj-cs.3259/fig-2

Table 4 Integrated framework linking system layers to ETA prediction.

System layer	Primary role in ETA prediction	Techniques	Key gains
Spatial temporal modeling	Improve accuracy using maps, traffic trends, long-term data, and engineered features.	Map refinement, GNNs, temporal embeddings, Long-Short Term Memory (LSTM)/autoencoders.	Reduced spatial error; stronger generalization.
Behavioral adaptation	Capture driver variability and enable real-time fleet recalibration.	Telematics, behavioral clustering, RL, multi-agent RL.	Personalized and dynamic ETAs.
Environmental robustness	Handle weather disruptions and rare anomalies.	Weather fusion, anomaly detection, ensemble models.	Robustness in volatile settings.
Feedback and fairness	Correct models <i>via</i> feedback and mitigate algorithmic bias.	Crowdsourcing, federated learning, bias audits.	Self-healing and equitable predictions.

with historical patterns to dynamically adjust speed predictions on affected routes. For instance, Uber's ETA models utilize weather overlays to predict delays due to snowfall, ensuring that travel times reflect the reality of adverse conditions. Studies have shown that snowstorms can reduce travel speeds in urban areas by up to 30%, while heavy rain can increase highway travel times by 15–20%.

- **Fog and reduced visibility** pose additional challenges. Fog obscures road signs, landmarks, and other vehicles, necessitating slower speeds and greater caution. Models incorporate geospatial data along with visibility indices to identify areas prone to fog-related delays. For example, routes traversing valleys or coastal regions with frequent fog are weighted differently in models compared to urban routes with clearer visibility. This differentiation helps systems better capture the variability in travel times caused by reduced visibility.
- **Localized weather patterns**, such as sudden thunderstorms or hail, introduce unique disruptions. These intense but short-lived events affect specific road segments, creating unpredictable slowdowns. Advanced ETA systems address this variability by integrating high-resolution, segment-level weather forecasts. For instance, DoorDash's ETA models leverage multi-task learning to combine real-time weather updates with traffic data, enabling accurate predictions for localized disruptions. For example, during a thunderstorm, segment-specific delays caused by localized flooding are incorporated into the model, which enables real-time rerouting.
- **Long-term seasonal effects**, such as recurring winter storms or monsoon seasons, also require careful consideration. Historical weather data is often used to train ML models to anticipate seasonal disruptions. For instance, in snowy regions, models account for typical snow-clearing times for major roads, ensuring ETAs are adjusted for delays caused by recurring winter conditions. Similarly, monsoon-prone regions benefit from

models trained to incorporate historical flooding patterns and road closures during peak rainfall periods. These adjustments allow ETA systems to remain robust during prolonged adverse conditions.

- **Cold-start effects after adverse weather events** also degrade ETA accuracy, particularly during the first trips of the day following snowfall, freezing rain, or low temperatures. In these conditions, vehicles may experience prolonged warm-up times, battery drain (especially in Electric vehicles (EVs)), and traction loss on untreated roads. These delays are often unaccounted for in standard ETA models, leading to systematic underestimation of travel times in the early morning. Moreover, secondary and residential streets may remain snow-covered long after main arteries are cleared, further compounding delay. Incorporating cold-start flags, weather-time interactions, and vehicle-level telemetry can help ETA systems dynamically adjust early-day predictions under wintry or icy conditions.
- **Quantifying uncertainty due to weather** remains essential for user trust and system reliability. Weather conditions are inherently unpredictable and vary widely across regions, which introduces significant variability into predictions. Probabilistic modeling approaches, such as those adopted by DoorDash and Shippeo, address this issue by quantifying uncertainty through confidence intervals or ranges for ETAs. For instance, during extreme weather events, systems may predict delays with a range (e.g., 15–30 min) rather than a fixed ETA, thereby empowering users to plan accordingly. Providing this transparency improves both user satisfaction and operational reliability.
- **Weather as a secondary variable** introduces indirect effects on ETA predictions, such as driver behavior and vehicle performance: —Driver behavior and comfort adjustments: Extreme temperatures prompt drivers to modify their vehicle’s internal environment. In hot conditions, air conditioning usage increases, while in freezing temperatures, heaters are used to maintain cabin comfort. These adjustments can increase fuel or battery consumption, indirectly influencing travel times. Furthermore, extreme weather conditions often lead to increased caution among drivers, who may adopt slower speeds or take additional rest breaks. —Vehicle efficiency under varying weather conditions: Weather impacts vehicle performance differently depending on the fuel source: *Biodiesel engines*: Pure biodiesel (B100) experiences reduced efficiency in freezing temperatures due to increased viscosity, thus necessitating preheating for optimal performance. Blends like B20, which include conventional diesel, mitigate these cold-weather issues and maintain better efficiency under such conditions. *EVs*: Extremely cold or hot temperatures negatively affect EV battery performance. Studies indicate that sub-zero temperatures can reduce EV battery range by 20–30%, because more energy is diverted to heating the cabin or maintaining battery temperature. Conversely, moderate temperatures improve efficiency, thereby enabling longer ranges. *Conventional fuel engines*: Gasoline and diesel engines consume more fuel during cold starts or in freezing conditions, when engines take longer to reach optimal operating temperatures. This is particularly impactful during short trips in winter regions.

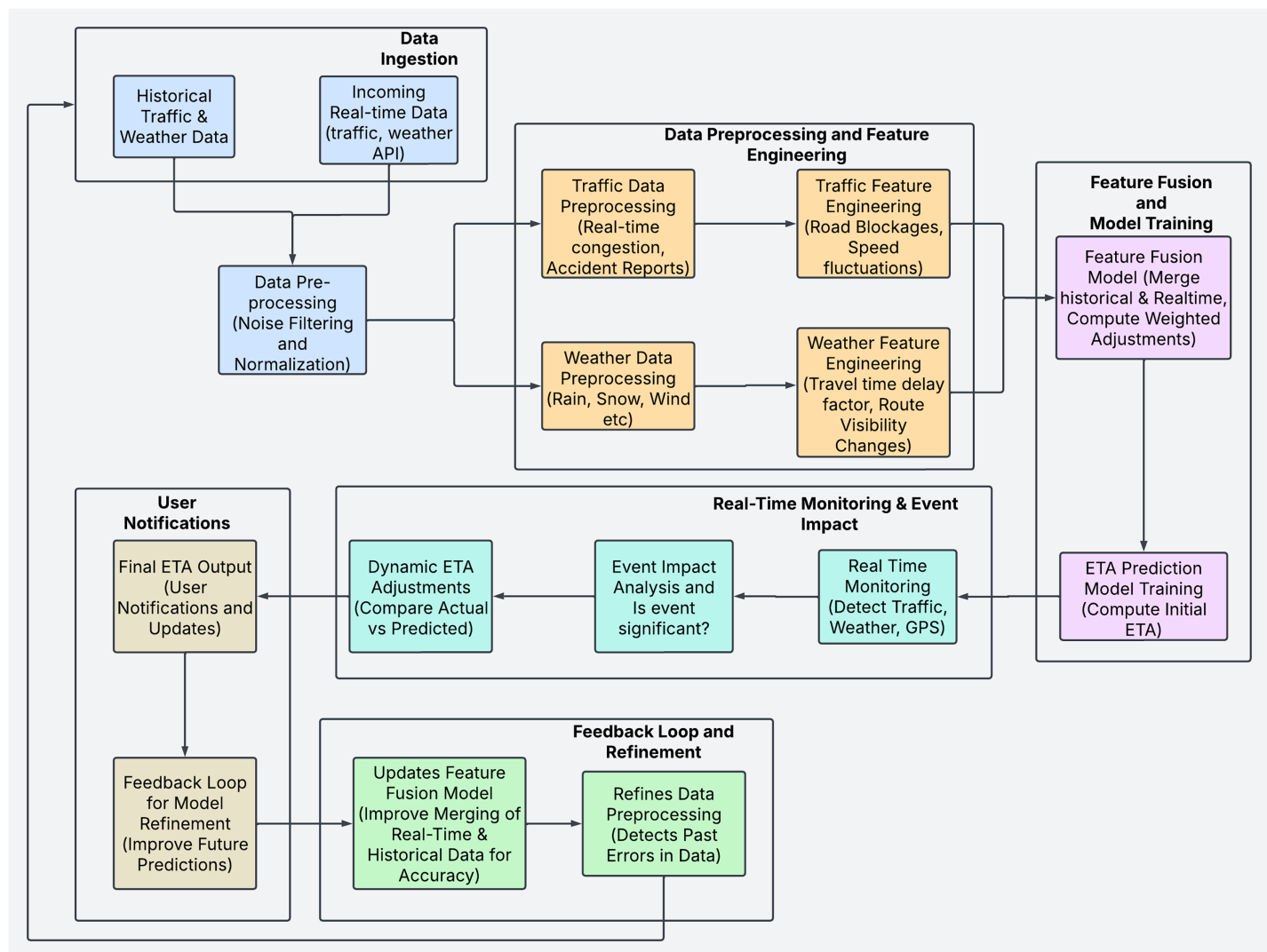


Figure 3 End-to-end ETA prediction system with real-time adjustments and feedback loop. This flowchart illustrates the complete workflow of an ETA prediction system, integrating historical and real-time data sources for accurate arrival time estimations. The process begins with data ingestion, where traffic, weather, and GPS data are collected and preprocessed separately. Feature engineering extracts relevant attributes before merging them in the feature fusion model. The ETA prediction model computes an initial estimate, which is continuously refined based on real-time event monitoring. If significant deviations occur, dynamic ETA adjustments are applied, and the final ETA is updated for user notifications. A feedback loop ensures continuous model refinement by improving feature fusion and preprocessing for future predictions, making the system adaptive and self-learning.

Full-size DOI: 10.7717/peerj-cs.3259/fig-3

- Advanced approaches to modeling weather effects** include probabilistic models, graph-based methods, and ensemble techniques. Probabilistic models quantify variability, while ensemble methods combine historical weather data with real-time updates to dynamically adapt predictions. Graph-based models, such as those used by Google Maps, incorporate weather as an edge attribute, enabling the system to predict delays and suggest reroutes during adverse conditions. Additionally, hyperlocal weather APIs, combined with spatiotemporal embeddings, improve regional predictions, ensuring that both short-term and localized disruptions are accounted for effectively.

Weather affects ETA predictions both directly—*via* traffic disruptions—and indirectly through driver behavior and vehicle performance. Integrating real-time weather data, historical patterns, and secondary factors improves prediction accuracy, user trust, and operational resilience across varied conditions.

TIP 6: LEVERAGING HISTORICAL DATA

Historical data plays a critical role in ETA prediction by enabling models to learn temporal patterns, account for route-specific congestion, and adapt to long-term infrastructure changes. By analyzing travel times across regions and timeframes, machine learning models can uncover recurring trends and anomalies to improve prediction accuracy. However, effective use of historical data requires addressing challenges such as temporal inconsistencies, missing values, spatial misalignments, and noise. Mitigating these issues is essential to ensure reliable, bias-free ETA forecasts.

- Temporal patterns challenges.** Historical data enables ETA models to capture long-term temporal patterns, such as rush-hour congestion, weekend traffic reductions, and seasonal variations. For instance, weekday morning commutes often experience travel times that are 20–30% higher than midday travel, while weekends generally exhibit smoother traffic conditions. By incorporating these temporal trends, models can dynamically adjust ETAs based on the time of day or season. This capability enables context-aware forecasting, leading to more precise predictions. However, these temporal patterns are often skewed by temporal imbalance, where models over-rely on high-traffic periods (*e.g.*, weekday commutes) and under-sample off-peak hours. This imbalance leads to systematic inaccuracies, particularly under underrepresented conditions, such as nighttime driving or holiday travel. Techniques such as data augmentation, transfer learning, and reinforcement learning can mitigate these biases, ensuring that models generalize effectively across all timeframes.
- Route-specific optimization and data challenges.** By analyzing historical data, ETA systems can identify recurring congestion patterns or bottlenecks that occur at specific times on certain road segments. For example, a road segment near a school may consistently experience delays during morning drop-off hours. Advanced systems, such as Uber’s DeepETA and Google Maps, leverage this segment-level granularity to refine predictions, achieving reductions in prediction error of up to 15% in congested urban areas. Despite these benefits, inconsistent spatial representations across different mapping providers pose a major obstacle to route-specific optimization. Variations in granularity, segment classifications, and topological inconsistencies can lead to discrepancies in travel time estimations. To address this, topological alignment methods and graph embedding techniques are used to harmonize disparate spatial data, thus ensuring a unified and reliable view of road networks (*Bast et al., 2016*).
- Model initialization and the impact of missing data.** Historical data plays a critical role in model initialization, particularly when assessing driver behavior early in a journey. At the beginning of a trip, real-time data may be limited or unavailable, which makes historical records essential for providing an initial baseline. For instance, if historical

data indicates that a particular driver tends to drive conservatively during morning routes, this information can be incorporated to improve early-stage ETA accuracy. However, missing data frequently arises due to network interruptions, sensor malfunctions, or GPS dropouts, particularly in remote areas or dense urban environments. These data gaps distort trajectory models and can significantly degrade ETA accuracy. Nearest-neighbor interpolation, and generative modeling have demonstrated up to 90% accuracy in reconstructing short-term missing values, thereby ensuring model continuity and robustness.

- **Long-term adaptability and noise challenges:** Historical data play a critical role in enabling ETA models to adapt to long-term changes in traffic patterns resulting from infrastructure development, road closures, and urban sprawl. For instance, the introduction of a new bypass or arterial route can significantly redistribute traffic flow, alleviating congestion on adjacent road segments. Empirical studies report that updating predictive models to reflect infrastructure changes can reduce ETA prediction errors by up to 20% in rapidly evolving urban regions.
- However, the utility of historical and real-time data is often compromised by noise introduced through sensor inaccuracies, timestamp mismatches, or urban signal interference. One common issue is GPS multipath distortion, where signals reflect off buildings and result in erroneous vehicle positioning—often placing vehicles on adjacent roads—thereby skewing travel-time calculations.
- To address these challenges, state-of-the-art ETA systems employ filtering and correction techniques such as Kalman filters and redundancy-based sensor fusion. These approaches smooth noisy observations and integrate data from multiple sources (e.g., GPS and inertial sensors) to improve positional accuracy and temporal alignment. Such enhancements are particularly valuable in dense urban settings where signal obstructions and high vehicle density exacerbate measurement noise, ultimately improving the robustness of ETA predictions.
- **Integrated framework for historical data utilization.** Effectively leveraging historical data requires a unified framework that integrates long-term patterns with real-time updates. Hybrid models, which combine spatiotemporal embeddings, imputation techniques, and noise filtering, enhance model reliability under diverse conditions. For instance, graph-based frameworks represent road networks as graphs, where intersections act as nodes and edges encode traffic flow, thereby enabling dynamic reconciliation of spatial and temporal inconsistencies. Additionally, edge computing processes localized corrections in real-time, reducing latency and ensuring that predictions remain responsive to evolving conditions. These advancements enable ETA models to dynamically integrate historical knowledge with real-time traffic signals, thereby bridging the gap between past insights and current road conditions.

Effectively integrating historical data with real-time insights is crucial for achieving high-precision ETA predictions. However, addressing challenges such as data sparsity, spatial inconsistencies, and temporal imbalance is equally important. Future

Table 5 Unexpected events and their impact on ETA predictions. This table outlines various disruptions, their effect on ETA accuracy, and mitigation strategies.

Event type	Impact on ETA	Mitigation strategy
Road accidents	Sudden traffic congestion, severe delays, and overflow into secondary roads.	Real-time traffic monitoring and dynamic rerouting using AI congestion models.
Flash floods	Road closures and detours due to inaccessible routes.	Integration of weather feeds with road data for proactive rerouting using ML optimization.
Public events	Localized congestion from marathons, parades, or concerts.	Use of event calendars and historical traffic to predict hotspots; apply temporary routing rules.
Strikes or protests	Unpredictable disruptions and law enforcement blockades.	Monitoring of news, social media, and police reports; route adaptation using AI-based threat models.
Extreme weather	Slower speeds and reduced visibility due to snow, fog, or storms.	High-resolution weather forecasting with vehicle sensor feedback to adjust ETA dynamically.
Infrastructure damage	Long detours due to bridge collapses or sinkholes.	Structural monitoring systems feed into dynamic routing to avoid affected zones.
Wildlife crossings	Slowdowns in rural areas from frequent animal movement.	Geofencing and seasonal data guide alerts and speed adjustments.

advancements in federated learning, Vehicle-to-Everything (V2X) communication, and real-time anomaly detection will continue to bridge the gap between historical trends and evolving traffic conditions, thereby enabling scalable, resilient, and adaptive ETA models.

TIP 7: ACCOUNTING FOR UNEXPECTED EVENTS IN ETA PREDICTIONS

Accurate ETA predictions require a balance between leveraging historical data to model long-term traffic patterns and dynamically adapting to unexpected events that deviate from established norms. Integrating these two approaches is essential for building robust and reliable ETA prediction systems, capable of addressing diverse real-world scenarios (Jiang et al., 2023; Yuan & Li, 2021; Wen et al., 2024) (See Table 5 regarding the effects of various unexpected events).

- **Historical data serves as a robust foundation for ETA predictions.** By capturing stable traffic patterns and long-term dynamics, historical data enables models to anticipate recurring phenomena such as diurnal congestion patterns, seasonal variations, and route-specific delays (Alhudhaif & Polat, 2024). For example, predictable trends, such as morning and evening rush hours or reduced congestion during midday, offer critical insights for refining model accuracy under routine conditions. Moreover, gradual changes in traffic dynamics—arising from urban expansion or new infrastructure—can be integrated into historical datasets to ensure predictions remain relevant. For instance, new bypass routes may reduce congestion on primary roads, thus necessitating updates to model parameters (Bast et al., 2016).
- **Unexpected large-scale disruptions, present significant challenges.** These anomalies deviate from routine traffic patterns and can drastically alter spatial and temporal dependencies. Examples include multi-vehicle collisions that cause highway closures,

flash floods that reroute traffic, or public events like protests that create bottlenecks (Santhosh, Dogra & Roy, 2020). Historical models often fail to anticipate such deviations, which leads to inaccuracies in ETA predictions. To mitigate these challenges, real-time data integration is crucial. Inputs from traffic sensors, crowdsourced reports, and live GPS updates enable modern systems like Waze and Google Maps to dynamically recalibrate predictions, adapting to disruptions as they unfold (De, 2022; Sasson, 2023).

- Long-tail events** represent another critical category of unexpected scenarios in ETA predictions. These events include rare and infrequent situations, such as deliveries to remote rural areas, operations during extreme weather conditions, or travel at atypical times, like late-night hours or holidays (Lu & Parekh, 2021). Long-tailed scenarios are often underrepresented in training datasets, making it difficult for models to effectively capture their unique characteristics (Zhang et al., 2023). One significant challenge with long-tail events is the variability of features specific to these scenarios. For instance, rural deliveries often involve longer travel distances, fewer alternative routes, and unpredictable road conditions (Shippeo, 2024). Similarly, extreme environmental conditions—such as hurricanes or snowstorms—introduce disruptions like reduced visibility, slower travel speeds, and prolonged delays (Huang, Wu & Lv, 2021). These factors deviate sharply from the structured, high-density patterns observed in urban environments, leading to less accurate predictions in these edge cases. Temporal sparsity further complicates long-tail events because reduced vehicle density or altered route availability during holidays or off-peak hours introduces distinct traffic patterns that are often overlooked during model training. To address long-tail events effectively, ETA systems must incorporate mechanisms such as data augmentation, domain adaptation, or ensemble approaches to improve generalization to these infrequent scenarios (Gal & Ghahramani, 2016). Without these solutions, systems risk failing to provide accurate ETAs when precision is most critical, such as during emergency logistics or severe weather conditions (Gupta, Gulla & Mancini, 2023).
- Hybrid frameworks are essential for effectively integrating historical and real-time data.** Techniques such as graph-based neural networks and spatiotemporal embeddings allow systems to encode spatial and temporal dependencies, enabling rapid updates to predictions (Battaglia et al., 2018). For instance, graph-based models employed by platforms like Google Maps analyze live traffic flows to provide real-time rerouting and accurate ETAs (De, 2022). Recurrent neural networks and long short-term memory networks further enhance a system's ability to capture sequential dependencies in historical data while dynamically adapting to real-time anomalies (Hochreiter, 1997). Additionally, ensemble methods combine historical trend models with anomaly detection modules, thereby ensuring robustness across diverse scenarios (Mondal, 2022).
- Event-based edge cases:** Public events such as marathons, parades, protests, and large festivals introduce complex, short-term disruptions that can render ETA models ineffective. These scenarios often involve planned road closures, restricted zones, and high pedestrian density—all of which may not be reflected in real-time traffic feeds or

historical patterns. Failure to account for such events can result in significant ETA errors and missed deliveries. Integrating city event calendars, crowd intelligence, and event-detection APIs can improve system responsiveness and prediction accuracy during these localized but high-impact disruptions.

- **Quantifying uncertainty is critical for improving user trust.** Unexpected events inherently introduce variability into ETA predictions. Probabilistic models address this by providing confidence intervals rather than single-point estimates. For example, during traffic disruptions, platforms like Uber and DoorDash communicate ETA ranges that account for potential delays, which allows users to set realistic expectations (*Jiang et al., 2024; Uber AI, Data/ML, 2022*). Such transparency is particularly crucial in high-stakes applications, such as logistics or emergency response, where timing accuracy is paramount.

In conclusion, balancing historical trends and unexpected events is vital for the development of reliable ETA prediction systems. By leveraging historical data for a stable predictive baseline, integrating real-time data for an adaptive response, and quantifying uncertainty to manage user expectations, modern systems achieve robustness and accuracy (*Naik, 2024; Gupta, Gulla & Mancini, 2023; Shippeo, 2024*). This dual approach ensures that ETA predictions remain effective across routine scenarios and rare disruptions, effectively addressing the dynamic demands of contemporary transportation networks.

TIP 8: LEVERAGING USER FEEDBACK IN ETA PREDICTIONS

ETA predictions are increasingly powered by models that synthesize real-time telemetry, historical traffic patterns, and algorithmic optimization. Yet, despite these advancements, predictive models often remain susceptible to bias, outdated assumptions, and failure to account for rare but high-impact disruptions. In this context, user feedback—gathered from drivers, passengers, and crowdsourced platforms—serves as a vital corrective mechanism. By incorporating real-world insights, ETA systems can iteratively recalibrate their forecasts, enhance local responsiveness, and improve accuracy over time (*Naik, 2024; Sasson, 2023*). Contemporary implementations increasingly integrate driver-reported anomalies, passenger feedback, and crowd-contributed traffic reports to strengthen model adaptability. When systematically processed, such feedback enables machine learning pipelines to more faithfully capture the nuances of complex transportation environments, thereby enhancing both the resilience and long-term precision of ETA predictions (*Jiang et al., 2023; Wen et al., 2024*).

- **Balancing accuracy and latency** is a key challenge when incorporating real-time feedback. ETA models must simultaneously achieve high precision and low-latency inference, particularly in applications such as ride dispatch and delivery routing. While high-precision models necessitate significant computation, real-world systems demand

sub-millisecond inference speeds. To optimize this trade-off, model compression techniques, such as *quantization*—which converts model weights to lower-precision formats—and *pruning*—which removes redundant neural connections—can significantly reduce inference times. For example, Uber’s DeepETA system reduced inference latency by 40% with minimal accuracy loss by using these optimizations (*Uber AI, Data/ML, 2022*). Another effective strategy is hybrid modeling. Companies like DoorDash use precomputed offline models, trained on historical data, to generate baseline ETAs, which are then refined in real time using lightweight, context-aware models that integrate driver feedback and live traffic conditions. In large-scale logistics operations, adaptive model selection further enhances scalability. For instance, Uber Freight employs dynamic model switching, choosing between high-precision and lightweight models based on traffic density, road complexity, and historical ETA accuracy (*Gupta, Gulla & Mancini, 2023*).

- **Ethical/privacy concerns** must be addressed when integrating user feedback into ETA models. Malicious or false reports can mislead systems if they are not properly validated. Also, privacy concerns arise from the continuous monitoring of driver and passenger locations. To mitigate these risks, companies employ data anonymization techniques and automated filtering mechanisms to ensure that feedback is both accurate and compliant with privacy regulations (*Bengio, Courville & Vincent, 2013*). Bias mitigation strategies, such as diversifying feedback sources and balancing training datasets, help to prevent models from overfitting to specific user-reported patterns (*Huang, Wu & Lv, 2021*).
- **Building user trust through transparent feedback handling** is equally important. A well-designed feedback system encourages more users to contribute valuable data. Many navigation platforms now provide real-time updates on how user feedback is applied. For instance, Waze updates road conditions within minutes of receiving validated driver reports, thereby enhancing trust in the system (*Sasson, 2023*). Additionally, modern ETA systems increasingly present confidence intervals rather than single-point estimates, which helps users anticipate variability in arrival times. Companies like DoorDash and Uber have adopted this approach, thus reducing user frustration when delays occur outside the initial estimate (*Zhang et al., 2024; Uber AI, Data/ML, 2022*). Finally, closing the feedback loop is crucial for long-term engagement. By notifying users that their input has influenced ETA predictions—e.g., “Your report helped refine predictions for this route”—companies reinforce the value of user participation and foster ongoing collaboration in improving ETA reliability.
- **User feedback provides critical real-time information** that enhances ETA predictions beyond what automated systems can detect. Drivers frequently encounter localized disruptions—such as temporary road closures, event-related congestion, and sudden construction zones—before such disruptions appear in map data. Uber’s DeepETA system leverages driver-reported alerts to dynamically adjust road network edge weights, prioritizing frequently flagged segments and ensuring more accurate route adjustments (*Uber AI, Data/ML, 2022*). Passengers also contribute valuable insights by identifying

systematic overestimations or underestimations in travel times. For example, Lyft clusters passenger complaints about ETA discrepancies by using spatial anomaly detection. This enables the system to identify geographical hotspots where predictions are frequently inaccurate. These insights are then used to prioritize model retraining efforts, thereby improving accuracy in problematic areas (Naik, 2024). To ensure effective feedback utilization, clustering algorithms group similar reports while anomaly detection filters false positives by comparing flagged issues against historical traffic and GPS data (Jiang et al., 2023).

- **User feedback is valuable in managing rare but high-impact disruptions**, such as major protests, severe weather, or public events, where historical data alone is insufficient. These anomalies introduce extreme variability in travel times, which requires dynamic adjustments. Google Maps integrates event detection APIs to preemptively recalibrate ETAs based on anticipated delays from large-scale events like concerts and sporting matches (De, 2022). Similarly, Uber Freight leverages anomaly detection models to identify sudden GPS velocity drops across multiple vehicles and enable real-time ETA adjustments to account for emerging disruptions (Gupta, Gulla & Mancini, 2023).

In conclusion, user feedback is essential for improving ETA predictions by capturing real-world anomalies, guiding retraining, and addressing rare disruptions. Platforms like Uber, Lyft, and Waze leverage driver alerts, passenger input, and dynamic retraining to build more adaptive systems. Techniques like confidence-weighted updates and anomaly detection help ensure scalability and accuracy, making real-time feedback a critical tool for trustworthy ETA models (Naik, 2024; Sasson, 2023).

TIP 9: OPTIMIZING FEATURE ENGINEERING FOR ACCURATE ETA PREDICTIONS

ETA prediction hinges on the selection and engineering of informative features that capture the complexities of travel dynamics. By integrating spatial, temporal, and contextual insights, models can enhance predictive precision and robustness against real-world uncertainties. Effective feature engineering not only improves model accuracy but also ensures adaptability across diverse transportation scenarios.

- **Spatial features: mapping the road network for better predictions.** A robust ETA model must account for road infrastructure, traffic flow patterns, and geographical constraints to enhance predictive reliability. (a) *Road Network Attributes*: The classification of roads (e.g., highways, arterial roads, and residential streets), lane capacity, toll structures, and elevation gradients significantly impact vehicle speed (Molinero, Murcio & Arcaute, 2017). For example, Waze's Smartsum model refines ETA predictions by incorporating directional constraints, intersection densities, and region-specific speed limits, thereby ensuring a context-aware approach to travel time estimation. (b) *Traffic Flow Patterns*: Real-time congestion data, when combined with historical traffic patterns, provides stronger predictive power than either source alone

([Ma et al., 2019](#)). For instance, DoorDash integrates live traffic feeds with historical route performance to anticipate fluctuations, particularly during peak hours, consequently improving ETA stability under varying demand conditions ([Jiang et al., 2024](#)).

(c) *Geospatial Embeddings*: Traditional road network models often struggle to capture regional traffic variations. To address this, Uber's DeepETA introduces geospatial embeddings, which adapt models to different travel environments, such as dense urban areas vs suburban or rural settings ([Uber AI, Data/ML, 2022](#)). These embeddings incorporate factors such as proximity to high-traffic zones, demographic movement patterns, and localized driver behavior, thus improving travel time estimation by leveraging deep representations of spatial context.

- **Temporal features: capturing the dynamic nature of traffic.** Traffic patterns are inherently dynamic and require temporal signals that enable models to adjust to real-time conditions and historical cyclic variations. (a) *Time of Day & Day of the Week*: Traffic congestion follows predictable cyclic patterns; rush hours, weekends, and nighttime travel all exhibit distinct flow characteristics. For example, Waze employs time-sensitive embeddings that dynamically modulate ETA predictions to reflect minute-level variations, effectively overcoming the limitations of static averaging techniques ([Petreanu, 2020](#)). (b) *Seasonality & Event-Based Disruptions*: Large-scale disruptions, such as holidays, major sporting events, and extreme weather conditions, introduce short-term anomalies that can significantly alter traffic flow. For instance, DoorDash integrates event-driven signals to incorporate these transient effects into its predictive framework, which allows for a more adaptive and context-aware ETA estimation ([Lu & Parekh, 2021](#)). (c) *Temporal Continuity*: The evolution of traffic conditions throughout a trip's duration necessitates a model's ability to track continuous changes. Monitoring consecutive time windows enables the differentiation between persistent trends (e.g., congestion buildup) and short-lived anomalies (e.g., temporary road closures). This temporal continuity is particularly valuable for long-haul travel predictions, where traffic conditions may evolve significantly during a single trip.
- **Advanced feature engineering techniques:** Advanced techniques further refine model inputs and allow for greater predictive accuracy by capturing complex dependencies in travel time estimation. (a) *High-Cardinality Feature Embeddings*: Many categorical variables—such as unique road segments, merchant locations, and transit hubs—require efficient encoding to avoid model inefficiencies. For example, DoorDash encodes store-specific preparation times into dense representations, enabling models to incorporate operational nuances without excessive computational overhead ([Zhang et al., 2024](#)). (c) *Feature Interactions & Nonlinear Relationships*: Certain feature interactions exhibit nontrivial dependencies that standard models may fail to capture. For instance, Uber leverages engineered cross-features, such as the interaction between traffic density and road type, to model congestion effects more accurately. For example, heavy congestion on a multi-lane highway may have a different impact on ETA compared to a similar level of congestion on a narrow urban street ([Gupta, Gulla & Mancini, 2023](#)).

Table 6 Summary of bias types in ETA prediction systems. The table highlights key sources of bias, their consequences, and actionable mitigation strategies.

Bias type	Challenges and consequences	Mitigation strategies
Geographic	Sparse rural data leads to poor generalization; models underperform in low-density areas.	Augment data using satellite imagery and synthetic data. Transfer learning from urban areas; crowdsource rural data.
Temporal	Fails during holidays, extreme weather, or off-peak patterns; ETAs become unreliable.	Time-weighted training, event-aware forecasting, and Bayesian uncertainty modeling.
Demographic and socioeconomic	Models favor affluent, well-connected regions; marginalized groups receive worse ETA quality.	Impose fairness constraints, audit bias with demographic metrics, and adjust loss functions.
Infrastructure	Outdated or missing road data hurts ETA accuracy in developing regions or rapidly changing areas.	Real-time updates from drivers and sensors; use online learning to adapt to infrastructure changes.
Driver behavior	Ignores variability in aggressive vs. cautious driving; personal ETAs become inaccurate.	Use driver embeddings, behavioral clustering, and RL-based real-time adaptation.
Weather and environmental	Limited extreme weather data leads to underestimated delays during storms or snowfall.	Incorporate live weather APIs; use probabilistic models and simulate adverse weather scenarios.
Feedback loop	Heavily used areas get more data, reinforcing neglect of less-traveled routes.	Dynamic sampling to rebalance training data; explore <i>via</i> RL; capture real-time user feedback.
Ethical and cultural	Uniform behavioral assumptions ignore local road customs, siesta hours, or etiquette.	Regional model customization; adaptive learning for cultural patterns; conduct fairness audits.

Feature engineering is crucial for building accurate and scalable ETA prediction models. Integrating spatial, temporal, and learned features improves robustness and reduces drift from real-time anomalies and latent factors. Companies like Uber, Waze, and DoorDash leverage structured embeddings and event-driven signals to enhance predictive accuracy. By optimizing geospatial and time-sensitive representations, organizations can create adaptive ETA models that remain reliable under dynamic road conditions.

TIP 10: ADDRESSING ALGORITHMIC BIAS IN ETA SYSTEMS

Algorithmic bias in ETA systems manifests when ML models produce systematically skewed predictions that disadvantage specific geographic, temporal, or demographic groups (Abdi & Amrit, 2021; Wen et al., 2024). Such biases not only erode user trust but also perpetuate inequities in service quality and operational efficiency. This section categorizes prevalent bias types, analyzes their societal impacts, and establishes a sequential auditing protocol to ensure equitable ETA predictions (See Table 6 for more details regarding the different biases and their effects.).

- Taxonomy and societal impacts:** —Bias in ETA systems arises from systemic gaps in data representation, model design, and operational feedback loops. Geographic bias occurs when sparse training data from rural or underserved regions leads to unreliable predictions, disproportionately affecting service reliability in these areas (Jiang et al., 2023). Studies might show that ETA accuracy in lower-income neighborhoods is 15–20% lower compared to affluent areas, reflecting this data sparsity. Similarly, temporal bias emerges from overemphasis on frequent patterns (e.g., weekday traffic) at the expense of rare events like holiday surges or extreme weather (Lu & Parekh, 2021). ETA systems trained on weekday data may exhibit 30% higher prediction errors during

late-night hours, disproportionately impacting shift workers. Demographic and socioeconomic biases, often rooted in historical data reflecting systemic inequalities, further marginalize low-income or minority communities (Huang, Wu & Lv, 2021). Wait times for ride-sharing in minority neighborhoods can be, on average, 25% longer during peak hours. Additional biases include infrastructure bias (incomplete road network data), driver behavior bias (overgeneralization of driving styles), and feedback loop bias, where underprediction in certain regions reduces future data collection, exacerbating neglect (Rahman, Abdel-Aty & Wu, 2021). Data collection rates can be up to 40% lower in underserved areas, creating negative feedback loops.

- Biased predictions amplify existing disparities. For instance, geographic bias limits emergency vehicle routing efficiency in underrepresented neighborhoods, while temporal bias compromises logistics during critical events like natural disasters (Lyu et al., 2019). Demographic biases reinforce socioeconomic divides by deprioritizing services to marginalized groups. Feedback loops compound these issues, as persistently inaccurate predictions in underserved areas deter user engagement, further limiting data collection (Shippeo, 2024).
- **Sequential auditing:** —A systematic audit of ETA systems requires sequential evaluation across six stages. (a) Define fairness criteria by establishing context-specific objectives, such as minimizing prediction error disparities between urban and rural zones (Yuan & Li, 2021). (b) Audit data representation by evaluating spatial, temporal, and demographic coverage. Heatmaps of trip density and statistical disparity tests (e.g., Kolmogorov-Smirnov for temporal distributions) reveal underrepresentation of rural routes or peak-hour edge cases (Pan et al., 2019). (c) Select quantifiable bias metrics, such as the disparate impact ratio (comparison of favorable prediction rates across groups) and group-specific mean absolute error (MAE) (Battaglia et al., 2018). (d) Analyze model behavior using explainability tools like SHAP (Shapley additive explanations) to identify features disproportionately influencing biased outcomes (Bengio, Courville & Vincent, 2013). For instance, ZIP code embeddings may exhibit outsized contributions to delays in low-income neighborhoods. (e) Validate scenario-specific performance through synthetic data injections (e.g., simulated traffic closures) & counterfactual queries (Che et al., 2018). (f) Implement continuous monitoring via fairness dashboards tracking demographic parity gaps & user feedback (De, 2022).
- **Sequential mitigation:** —Mitigation strategies span interdisciplinary collaboration. (a) Data-centric approaches include geospatial oversampling, where underrepresented regions are enriched with satellite-derived traffic patterns or crowdsourced GPS traces (Niu & Silva, 2020). Temporal gaps are addressed through synthetic event generation, such as simulating traffic patterns during cultural festivals using agent-based models (Gal & Ghahramani, 2016). (b) Model-centric solutions involve fairness-aware regularization, where loss functions penalize demographic or geographic error disparities (Jiang et al., 2023). (c) Operational strategies emphasize dynamic recalibration using Internet of Things (IoT) sensor data and real-time map correction

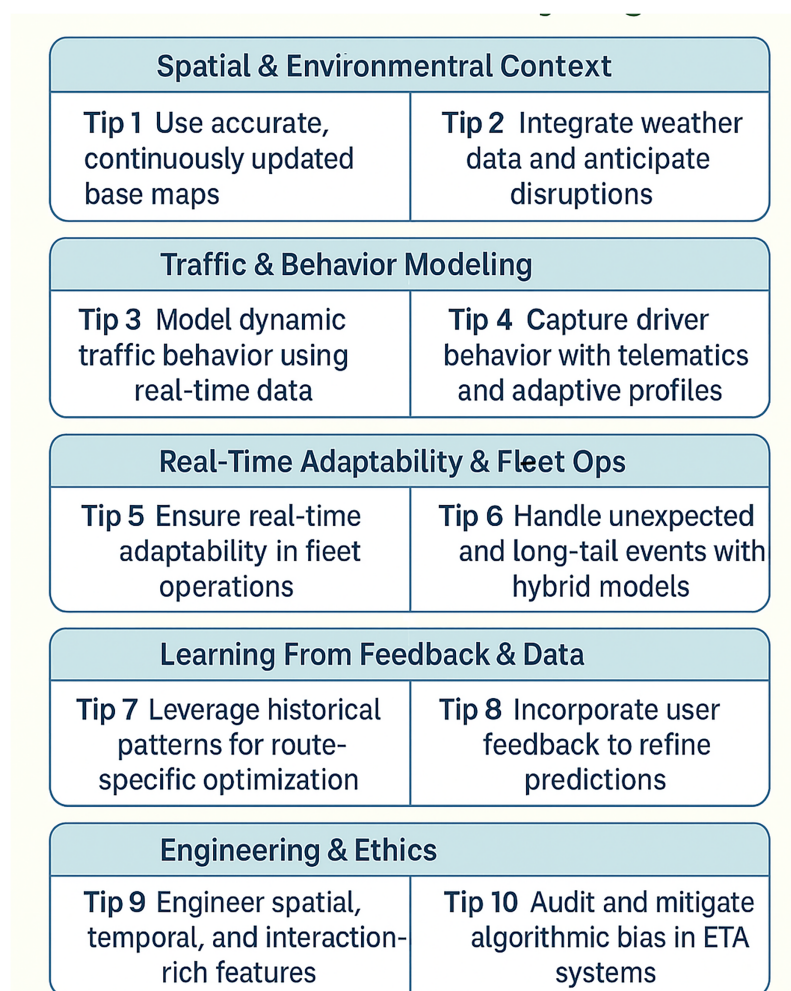


Figure 4 Schematic overview of the ten evidence-based strategies for improving ETA predictions in delivery logistics. The tips are grouped into five thematic domains—(1) spatial & environmental context, (2) traffic & driver-behavior modeling, (3) real-time fleet adaptability, (4) learning from feedback & historical data, and (5) engineering & ethics—highlighting how complementary data sources and model practices collectively boost accuracy, robustness, and fairness.

Full-size DOI: 10.7717/peerj-cs.3259/fig-4

pipelines. For example, systems like Baidu’s DuARE incorporate aerial imagery and vehicle trajectory data to continuously extract and update road networks in underrepresented or changing regions (Yang et al., 2022). This enables ETA models to better reflect real-world infrastructure changes and reduce geographic bias. In addition, transparency protocols—such as public fairness dashboards and periodic audit disclosures—help build user trust and institutional accountability.

In conclusion, algorithmic bias in ETA systems is not merely a technical flaw—it is a structural risk that deepens existing social inequities. Ensuring fairness requires a holistic strategy: rigorous bias audits, equitable data practices, model-level corrections, and operational transparency (Liu et al., 2024). By embedding fairness into every layer—from

data collection to deployment—ETA systems can become not only more accurate, but also more just, inclusive, and trustworthy in real-world transportation networks (Guo *et al.*, 2019).

FUTURE DIRECTIONS

Several research priorities can strengthen the next generation of ETA prediction systems. First, there is an urgent need for standardized benchmarking frameworks. Despite rapid model innovation, the lack of open datasets and unified evaluation metrics limits direct comparison. Future work should focus on creating public, multimodal ETA datasets and performance leaderboards to enable reproducible model evaluation across different geographic and operational contexts. Second, future reviews should adopt transparent synthesis methodologies, such as PRISMA-style frameworks, to improve replicability. This includes reporting specific search strings, screening processes, and inclusion/exclusion criteria, supported by flow diagrams or numerical summaries. Third, explainability and interpretability must move beyond raw accuracy. Lightweight explainable AI that expose feature importance and uncertainty estimates will be essential for regulatory oversight and operational trust, especially in safety-critical environments. Finally, fairness and privacy require urgent attention. Algorithmic bias—geographic, temporal, or demographic—must be addressed through structured audits and mitigation strategies. Federated learning offers a path toward privacy-preserving collaboration, especially for cross-fleet or cross-regional model development. By advancing these directions—benchmarking, transparency, explainability, and equity—the field can build scalable, ethical, and trustworthy ETA systems for global deployment.

CONCLUSION

This review distilled a decade of research into ten actionable strategies to transform ETA prediction from static point estimates into dynamic, context-aware systems. By integrating advances in machine learning, spatiotemporal modeling, traffic physics, and real-time data ingestion, we outlined how modern ETA frameworks can boost accuracy, reduce latency, and remain resilient under volatile real-world conditions. However, technological progress alone is not enough. The field still faces persistent limitations—model opacity, data sparsity in underrepresented regions, vulnerability to disruptions, and algorithmic bias—that constrain large-scale deployment and erode user trust. Addressing these challenges will require a shift toward more transparent, interpretable, and equitable system design.

To this end, our ten tips are not isolated recommendations but interconnected strategies. As illustrated in Fig. 4, these tips naturally cluster into five foundational domains: Spatial & Environmental Context, Traffic & Driver-Behavior Modeling, Real-Time Fleet Adaptability, Learning from Feedback & Historical Data, Engineering & Ethics. This framework highlights how diverse modeling techniques, data sources, and deployment practices can work in tandem to improve ETA system performance across operational, technical, and societal dimensions. Ultimately, realizing robust and fair ETA predictions will demand sustained efforts across algorithm design, infrastructure modernization, and governance. Yet the foundations are now in place. With careful

engineering and principled development, the next generation of ETA systems can deliver not only operational efficiency for fleets and logistics platforms—but also safer, more equitable mobility for users worldwide.

ACKNOWLEDGEMENTS

The authors acknowledge the use of an AI tool, specifically ChatGPT, to support language refinement and content organization. Final decisions and all intellectual contributions remain those of the author.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this works.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Aasim Ayaz Wani conceived and designed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

This is a literature review.

REFERENCES

- Abdi A, Amrit C. 2021.** A review of travel and arrival-time prediction methods on road networks: classification, challenges and opportunities. *PeerJ Computer Science* **7**(4):e689 DOI [10.7717/peerj-cs.689](https://doi.org/10.7717/peerj-cs.689).
- Alhudhaif A, Polat K. 2024.** Spatio-temporal characterisation and compensation method based on CNN and LSTM for residential travel data. *PeerJ Computer Science* **10**(7):e2035 DOI [10.7717/peerj-cs.2035](https://doi.org/10.7717/peerj-cs.2035).
- Amini RE, Al Haddad C, Batabyal D, Gkena I, De Vos B, Cuenen A, Brijs T, Antoniou C. 2023.** Driver distraction and in-vehicle interventions: a driving simulator study on visual attention and driving performance. *Accident Analysis & Prevention* **191**:107195 DOI [10.1016/j.aap.2023.107195](https://doi.org/10.1016/j.aap.2023.107195).
- Bast H, Delling D, Goldberg A, Müller-Hannemann M, Pajor T, Sanders P, Wagner D, Werneck RF. 2016.** Route planning in transportation networks. *Algorithm Engineering: Selected Results and Surveys* **9220**:19–80 DOI [10.1007/978-3-319-49487-6_2](https://doi.org/10.1007/978-3-319-49487-6_2).
- Battaglia P, Hamrick JBC, Bapst V, Sanchez A, Zambaldi V, Malinowski M, Tacchetti A, Raposo D, Santoro A, Faulkner R, Gulcehre C, Song F, Ballard A, Gilmer J, Dahl GE, Vaswani A, Allen K, Nash C, Langston VJ, Dyer C, Heess N, Wierstra D, Kohli P, Botvinick M, Vinyals O, Li Y, Pascanu R. 2018.** Relational inductive biases, deep learning, and graph networks. ArXiv DOI [10.48550/arXiv.1806.01261](https://doi.org/10.48550/arXiv.1806.01261).

- Bengio Y, Courville A, Vincent P. 2013.** Representation learning: a review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(8):1798–1828 DOI 10.1109/tpami.2013.50.
- Che Z, Purushotham S, Cho K, Sontag D, Liu Y. 2018.** Recurrent neural networks for multivariate time series with missing values. *Scientific Reports* 8(1):6085 DOI 10.1038/s41598-018-24271-9.
- Chiang Y-Y, Leyk S, Knoblock CA. 2014.** A survey of digital map processing techniques. *ACM Computing Surveys (CSUR)* 47(1):1–44 DOI 10.1145/2557423.
- Chowdhury M, Dey K, Apon A. 2024.** *Data analytics for intelligent transportation systems*. Amsterdam: Elsevier.
- De S. 2022.** Demystifying eta prediction in google maps. Available at <https://medium.com/@srijanak2001/demystifying-eta-prediction-in-google-maps-ab2a63ce5cff> (accessed 4 December 2024).
- Evmides N, Aslam S, Ramez TT, Michaelides MP, Herodotou H. 2024.** Enhancing prediction accuracy of vessel arrival times using machine learning. *Journal of Marine Science & Engineering* 12(8):1362 DOI 10.3390/jmse12081362.
- Gal Y, Ghahramani Z. 2016.** Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: *International Conference on Machine Learning*. PMLR, 1050–1059.
- Guo S, Lin Y, Feng N, Song C, Wan H. 2019.** Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence* 33(1):922–929 DOI 10.1609/aaai.v33i01.3301922.
- Gupta M, Gulla M, Mancini A. 2023.** A look inside the AI engine powering on-time arrivals—Uber freight. Available at <https://www.uberfreight.com/blog/estimating-shipment-arrival-ai/> (accessed 4 December 2024).
- He X, Pan J, Jin O, Xu T, Liu B, Xu T, Shi Y, Atallah A, Herbrich R, Bowers S, Candela JQ. 2014.** Practical lessons from predicting clicks on ads at Facebook. In: *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising, ADKDD’14*. New York, NY, USA: Association for Computing Machinery, 1–9.
- Hochreiter S. 1997.** Long short-term memory. In: *Neural Computation*. Cambridge: MIT-Press.
- Huang Z, Wu J, Lv C. 2021.** Driving behavior modeling using naturalistic human driving data with inverse reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems* 23(8):10239–10251 DOI 10.1109/tits.2021.3088935.
- Jiang W, Luo J, He M, Gu W. 2023.** Graph neural network for traffic forecasting: the research progress. *ISPRS International Journal of Geo-Information* 12(3):100 DOI 10.3390/ijgi12030100.
- Jiang Z, Xu Q, Zhang C, Jenq H, Luo J. 2024.** Precision in motion: deep learning for smarter eta predictions—DoorDash. Available at <https://careersatdoordash.com/blog/deep-learning-for-smarter-eta-predictions/> (accessed 4 December 2024).
- Kumar P, Gupta GP, Tripathi R, Garg S, Hassan MM. 2021.** DLTIF: deep learning-driven cyber threat intelligence modeling and identification framework in IoT-enabled maritime transportation systems. *IEEE Transactions on Intelligent Transportation Systems* 24(2):2472–2481 DOI 10.1109/tits.2021.3122368.
- Li T, He Z, Wu Y. 2022.** An integrated route planning approach for driverless vehicle delivery system. *PeerJ Computer Science* 8(4):e1170 DOI 10.7717/peerj-cs.1170.
- Li Y, Yu R, Shahabi C, Liu Y. 2017.** Diffusion convolutional recurrent neural network: data-driven traffic forecasting. ArXiv DOI 10.48550/arXiv.1707.01926.

- Liu Y, Shen G, Liu N, Han X, Xu Z, Zhou J, Kong X. 2024. Traffic data imputation via knowledge graph-enhanced generative adversarial network. *PeerJ Computer Science* 10(4):e2408 DOI 10.7717/peerj-cs.2408.
- Lu D, Parekh P. 2021. Improving eta prediction accuracy for long-tail events—DoorDash. Available at <https://careersatdoordash.com/blog/improving-eta-prediction-accuracy-for-long-tail-events/> (accessed 4 December 2024).
- Lyu Y, Siddique A, Majid SH, Biglarbegian M, Gadsden S, Mahmud S. 2019. Electric vehicle battery thermal management system with thermoelectric cooling. *Energy Reports* 5(3):822–827 DOI 10.1016/j.egy.2019.06.016.
- Ma J, Chan J, Ristanoski G, Rajasegarar S, Leckie C. 2019. Bus travel time prediction with real-time traffic information. *Transportation Research Part C: Emerging Technologies* 105:536–549 DOI 10.1016/j.trc.2019.06.008.
- Molinero C, Murcio R, Arcaute E. 2017. The angular nature of road networks. *Scientific Reports* 7(1):4312 DOI 10.1038/s41598-017-04477-z.
- Mondal A. 2022. ML system design—eta prediction. Available at <https://mecha-mind.medium.com/ml-system-design-eta-prediction-9dc8000fd86b> (accessed 4 December 2024).
- Munigety CR, Mathew TV. 2016. Towards behavioral modeling of drivers in mixed traffic conditions. *Transportation in Developing Economies* 2(1):1–20 DOI 10.1007/s40890-016-0012-y.
- Naik R. 2024. Eta (estimated time of arrival) reliability at Lyft. Available at <https://eng.lyft.com/eta-estimated-time-of-arrival-reliability-at-lyft-d4ca2720bda8?gi=481612ab769f> (accessed 4 December 2024).
- Niu H, Silva EA. 2020. Crowdsourced data mining for urban activity: review of data sources, applications, and methods. *Journal of Urban Planning and Development* 146(2):4020007 DOI 10.1061/(asce)up.1943-5444.0000566.
- O’Neill S. 2022. How the Zoox robotaxi predicts everything, everywhere, all at once—Amazon science. Available at <https://www.amazon.science/latest-news/how-the-zoox-robotaxi-predicts-everything-everywhere-all-at-once> (accessed 4 December 2024).
- Pan Z, Liang Y, Wang W, Yu Y, Zheng Y, Zhang J. 2019. Urban traffic prediction from spatio-temporal data using deep meta learning. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York: ACM, 1720–1730.
- Petreanu Y. 2020. Under the hood: real-time eta and how Waze knows you’re on the fastest route. Available at <https://medium.com/waze/under-the-hood-real-time-eta-and-how-waze-knows-youre-on-the-fastest-route-78d63c158b90> (accessed 4 December 2024).
- Qi W, Shen Z-JM. 2019. A smart-city scope of operations management. *Production and Operations Management* 28(2):393–406 DOI 10.1111/poms.12928.
- Rahman MH, Abdel-Aty M, Wu Y. 2021. A multi-vehicle communication system to assess the safety and mobility of connected and automated vehicles. *Transportation Research Part C: Emerging Technologies* 124:102887 DOI 10.1016/j.trc.2020.102887.
- Santhosh KK, Dogra DP, Roy PP. 2020. Anomaly detection in road traffic using visual surveillance: a survey. *ACM Computing Surveys (CSUR)* 53(6):1–26 DOI 10.1145/3417989.
- Sasson R. 2023. Eta prediction at Waze using deep learning models. Available at <https://www.datascienceeconomics.com/post/eta-prediction-at-waze-using-deep-learning-models> (accessed 4 December 2024).
- Shi W, Zhang J, Zhong X, Chen X, Ye X. 2024. DTS-AdapSTNet: an adaptive spatiotemporal neural networks for traffic prediction with multi-graph fusion. *PeerJ Computer Science* 10(5):e2527 DOI 10.7717/peerj-cs.2527.

- Shippeo. 2024.** Improving eta predictions in transport: challenges and solutions. Available at <https://www.shippeo.com/blog/improving-eta-predictions-in-transport-challenges-and-solutions> (accessed 4 December 2024).
- Singh H, Kathuria A. 2021.** Analyzing driver behavior under naturalistic driving conditions: a review. *Accident Analysis & Prevention* **150**:105908 DOI 10.1016/j.aap.2020.105908.
- Szumaska EM, Jurecki R. 2020.** The effect of aggressive driving on vehicle parameters. *Energies* **13**(24):6675 DOI 10.3390/en13246675.
- Uber AI, Data/ML. 2022.** DeepETA: how Uber predicts arrival times using deep learning. Available at <https://www.uber.com/blog/deepeta-how-uber-predicts-arrival-times/> (accessed 4 December 2024).
- Wang Z, Fu K, Ye J. 2018.** Learning to estimate the travel time. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York: ACM, 858–866.
- Wen H, Lin Y, Wu L, Mao X, Cai T, Hou Y, Guo S, Liang Y, Jin G, Zhao Y, Zimmermann R, Ye J, Wan H. 2024.** A survey on service route and time prediction in instant delivery: taxonomy, progress, and prospects. *IEEE Transactions on Knowledge and Data Engineering* **36**(12):7516–7535 DOI 10.1109/tkde.2024.3441309.
- Yang J, Ye X, Wu B, Gu Y, Wang Z, Xia D, Huang J. 2022.** DuARE: automatic road extraction with aerial images and trajectory data at Baidu maps. In: *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. New York: ACM, 4321–4331.
- Yuan H, Li G. 2021.** A survey of traffic prediction: from spatio-temporal data to intelligent transportation. *Data Science and Engineering* **6**:63–85 DOI 10.1007/s41019-020-00151-z.
- Zhang Y, Kang B, Hooi B, Yan S, Feng J. 2023.** Deep long-tailed learning: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**(9):10795–10816 DOI 10.1109/tpami.2023.3268118.
- Zhang C, Warne L, Jiang Z, Xu Q, Jenq H. 2024.** Improving etas with multi-task models, deep learning, and probabilistic forecasts. Available at <https://careersatdoordash.com/blog/improving-etis-with-multi-task-models-deep-learning-and-probabilistic-forecasts/> (accessed 4 December 2024).