future transportation

MDPI

Review

Context-Aware Travel Time Prediction and Route Optimization Using Heterogeneous Traffic and Event Data: A Comprehensive Survey

Gianpaolo Ghiani [1], Emanuele Manni [1], Valentino Moretto [1], Sandra De Iaco [2,*], Monica Palma [2] and Gianluca Romano [2]

[1] Dipartimento di Ingegneria dell'Innovazione, Università del Salento, Via per Monteroni, 73100 Lecce, Italy; gianpaolo.ghiani@unisalento.it (G.G.); emanuele.manni@unisalento.it (E.M.)
[2] Dipartimento di Scienze dell'Economia, Università del Salento, Via per Monteroni, 73100 Lecce, Italy; monica.palma@unisalento.it (M.P.)
* Correspondence: sandra.deiaco@unisalento.it

Abstract

Real-time navigation systems are increasingly used to provide optimal driving routes together with accurate travel time predictions that reflect dynamic urban traffic conditions. Recent advances have focused on integrating structured traffic data from traditional APIs with unstructured, context-rich information extracted via semantic crawling of news websites and social media platforms. This survey reviews state-of-the-art approaches that combine these heterogeneous data sources to improve route planning and travel time estimation, with special attention to the challenges posed by incident detection, event extraction, and multimodal data fusion. We discuss core methodologies including natural language processing techniques for event recognition, machine learning models for traffic prediction, and graph-based routing algorithms, highlighting their advantages and limitations. Finally, we outline open research directions for building context-aware navigation systems able to adapt to real urban mobility conditions.

Keywords: context-aware navigation systems; real-time traffic prediction; time-dependent vehicle routing; graph neural networks; social media mining; natural language processing; systematic literature review

Check for updates

Academic Editors: Tianwen Zhang and Jaeyoung Lee

Received: 9 April 2026
Revised: 6 May 2026
Accepted: 18 May 2026
Published: 29 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The rapid expansion of urban populations and vehicle density has made real-time, traffic-aware navigation systems indispensable for modern mobility management. Urban mobility demand has increased by more than 30% in the last two decades worldwide, placing growing pressure on road infrastructures [1]. Modern navigation platforms are no longer expected to deliver only the shortest path; they must also provide reliable travel time predictions that adapt to highly dynamic conditions such as congestion, accidents, extreme weather, and socio-political disruptions. Recent studies emphasize that unreliable travel time predictions frustrate users and also lead to systemic inefficiencies in freight logistics and emergency response [2,3].

Traditional services such as Google Maps (https://www.google.com/maps, accessed on 17 May 2026), HERE (https://www.here.com, accessed on 17 May 2026), or TomTom (https://www.tomtom.com, accessed on 17 May 2026) rely heavily on structured API data to report congestion levels and incident updates. Only recently have some navigation

systems and apps attempted to incorporate events such as roadblocks, strikes, concerts, sports games, political rallies, and other local disruptions to adjust Expected Time of Arrival (ETA) predictions [4]. However, most of these systems are limited to issuing delay warnings near event locations—such as stadiums during games or large concerts. Notably, crowdsourced navigation platforms like Waze (https://www.waze.com, accessed on 17 May 2026) showcase the potential benefits of integrating user-generated content, but they remain constrained by participation bias and inconsistent data quality [5]. To address these shortcomings, recent research has focused on exploiting semantic crawling and event extraction techniques to transform noisy, unstructured textual streams from the web and social media into usable information [6]. These approaches detect and classify real-world events in real time and estimate their potential impact on traffic conditions. This supports more accurate travel time estimates and dynamic adjustments to driving directions.

Unlike existing surveys that focus separately on routing algorithms, traffic prediction [7,8], or social media mining for transportation [9], this work makes the following contributions:

1. *Unified scope.* It jointly addresses time-dependent routing, travel time prediction using both structured and unstructured data, and NLP-driven event extraction within a single review. To the best of our knowledge, no previous survey has covered these three areas together.
2. *End-to-end pipeline perspective.* Rather than treating each component in isolation, this review follows the data flow from acquisition (traffic APIs, web and social media crawling) through event extraction and predictive modelling to context-aware route optimization, highlighting how these stages interact and depend on one another.
3. *Cross-disciplinary bridge.* The surveyed literature spans three research communities: operations research (routing and shortest-path algorithms), machine learning (traffic forecasting models), and natural language processing (event detection from text). These communities typically publish in separate venues and rarely cite one another. By synthesizing their contributions, this review offers a reference point accessible to researchers in all three fields.
4. *Critical analysis of deployment challenges.* Beyond methodological advances, this survey examines practical barriers to real-world adoption, including data latency, event severity quantification, cross-city transferability, ethical concerns related to social media mining, and compliance with data protection regulations.

In this survey we use "context-aware" in a narrow, event-driven sense, consistent with the PRISMA search protocol described in Section 2. A navigation or travel time prediction system is context-aware when it integrates information about discrete events (accidents, road closures, weather disruptions, public gatherings, socio-political disturbances) with the more traditional inputs of historical traffic data and real-time sensor or API readings. Event information is extracted from structured or unstructured sources, typically news outlets and social media. We use "event severity" for the magnitude of a detected event's impact on travel times, from a local delay to wider congestion propagation on the network. Other forms of context-awareness relevant to navigation, including road geometry (grade, curvature), dynamic tolls, air quality, driving style, driver fatigue, point-of-interest (POI) data, and multimodal transfer planning, were not part of the thematic queries defined in Section 2 and therefore fall outside the scope of this review. They represent complementary research directions.

The remainder of this paper is organized as follows. After presenting the methodology used to select, screen, analyse and classify the various referenced papers (Section 2), in Section 3 we describe how the problem of determining point-to-point routing directions is modelled in the literature and report the most relevant contributions. In Section 4 we give

an overview of the usage of the main commercial and governmental traffic data APIs in enabling real-time traffic analysis, followed by a description of the principal approaches to travel time prediction. Then, in Section 5 we review web and social media crawling techniques for extracting traffic-related events, followed by an analysis of travel time prediction methods that integrate such event-based data with traditional sources. Finally, some conclusions and suggestions for future research directions are reported in Section 6.

2. Research Methodology

The literature reviewed in this paper was identified following the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analysis) 2020 guidelines [10]. A systematic search was conducted across three academic databases (Scopus, Web of Science, and IEEE Xplore), complemented by Google Scholar for supplementary identification. Four search queries were designed to cover the three thematic areas of this review:

- *Query A* (time-dependent routing) defined by combining terms such as "time-dependent", "time-varying", "shortest path", "vehicle routing", "travel time", and "road network";
- *Query B* (traffic prediction) defined by combining the terms "traffic flow prediction", "traffic forecasting", "travel time prediction" with "deep learning", "neural network", "graph neural", "LSTM", "spatio-temporal";
- *Query C1* (social media and traffic) fixed by combining "social media", "Twitter", "Waze", "crowdsourcing" with "traffic incident", "traffic accident", "traffic congestion";
- *Query C2* (natural language processing-NLP and traffic) defined by combining "NLP", "natural language processing", "text mining", "geoparsing" with "road traffic", "traffic accident", "traffic event".

For full reproducibility, the Boolean expressions adopted in Scopus (and adapted to the syntax of each other database while preserving the same terms and logical operators) are reported in Appendix A.

All database queries were restricted to papers published during the period 2000–2025. Citation tracking was not subject to date restrictions, allowing the inclusion of seminal earlier works in the field. The search yielded a total of 23,094 records (16,593 from Scopus, 5589 from Web of Science, and 912 from IEEE Xplore). After removing 6098 duplicates, the remaining 16,996 records were screened by title and abstract against predefined inclusion and exclusion criteria. In particular, to efficiently handle the large volume of data, the screening phase was facilitated by automated keyword-based filtering.

This automated pre-screening step is explicitly provisioned by the PRISMA 2020 statement [10], whose standard flow diagram includes a dedicated field ("Records marked as ineligible by automation tools") for exactly this type of workflow; manual eligibility assessment was then performed by the authors on all candidate records that passed the automated filter, as detailed below.

Records were excluded if they were unrelated to transportation networks, not written in English, or outside the three thematic areas above defined by the search queries. This process resulted in the exclusion of 9646 records. The resulting 7350 candidate records were then sought for retrieval and assessed for eligibility by the authors through a sequential evaluation of titles, abstracts, and full texts. An additional 7 papers, not indexed in the above databases, were identified through Google Scholar and citation tracking. A total of 86 studies, directly addressing the three thematic areas defined by the search queries, were included in the final review. Beyond these, the reference list cites additional sources for methodological and contextual support that were not part of the systematic selection process. The complete PRISMA flow diagram is reported in Figure 1.

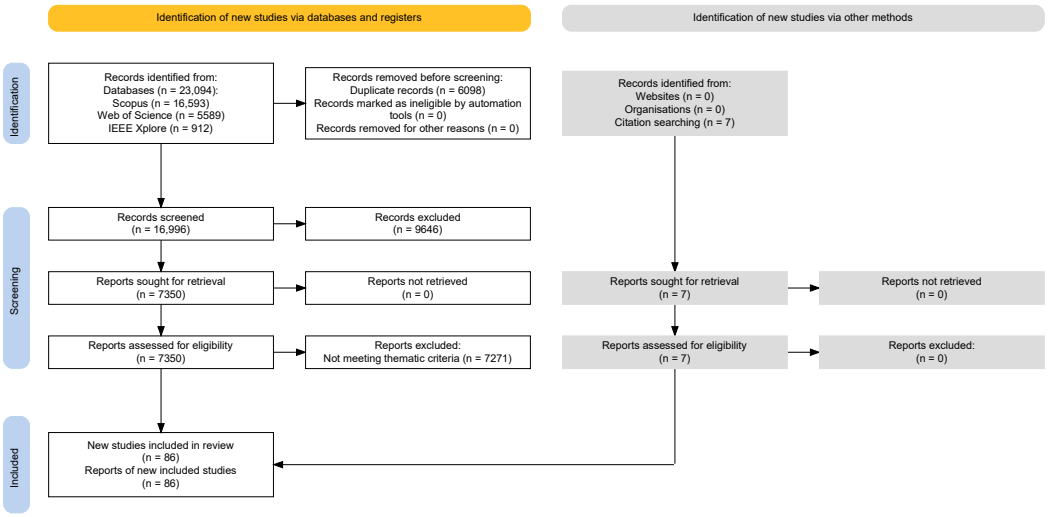

Figure 1. PRISMA 2020 flow diagram of the systematic literature search process.

The reference list of this paper cites 89 works in total: the 86 studies selected through the systematic search above, plus three foundational methodological references that are not subject to the thematic search—the PRISMA 2020 statement [10], the K-means clustering algorithm [11], and Affinity Propagation [12]. Among the 86 thematic citations, five references were added during the peer-review process on explicit reviewer suggestion, as authorised by the MDPI editorial policy: Cai et al. (2025) [13], Zhang et al. (2025) [14], and Lin et al. (2025) [15] after the first review round; Yuan and Li (2021) [16] and Pan et al. (2025) [17] after the second review round. These additions are cited as complementary supporting studies and do not alter the outcome of the PRISMA-based selection.

3. Computing Driving Directions

Computing driving directions entails efficiently identifying optimal point-to-point paths within large-scale, time-dependent transportation networks. These networks are characterized by dynamic factors such as fluctuating traffic conditions, road closures, and varying speed limits, which impact travel times and routing decisions. Depending on the specific application or user preferences, the optimization objective may vary. Common goals include minimizing overall travel duration to ensure timely arrivals, reducing fuel consumption to lower operational costs, or decreasing CO_2 emissions to promote environmentally sustainable travel. In many cases, modern navigation systems seek to balance multiple criteria simultaneously, requiring sophisticated multi-objective optimization techniques that account for trade-offs between efficiency, cost, and environmental impact.

Ideally, traffic forecasts for road travel times should integrate a wide range of information sources to improve accuracy and responsiveness. This includes historical congestion patterns that capture typical traffic behaviors, real-time sensor and GPS data reflecting current conditions, and unstructured data such as social media reports and event notifications that may indicate incidents or disruptions. Incorporating these heterogeneous data streams allows routing algorithms to generate more reliable and context-aware travel time predictions. Figure 2 displays an effective scheme of a travel time prediction and driving direction system based on structured traffic information and unstructured web/social media data.

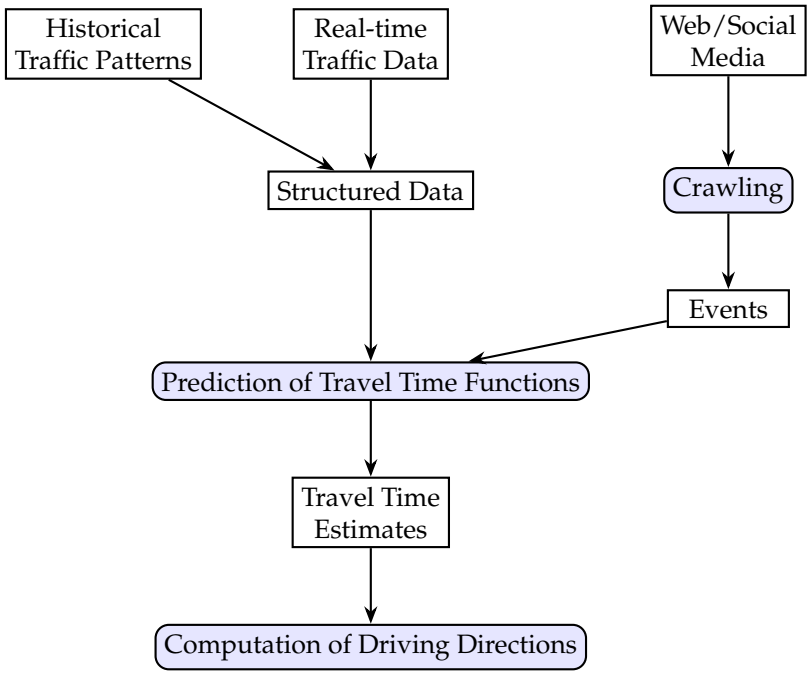

Figure 2. Block diagram of the travel time prediction and driving direction system. Structured inputs (historical and real-time traffic data) and unstructured inputs (web and social media content, processed through crawling and event extraction) feed the prediction module, which outputs travel time estimates used by the route optimization stage.

As illustrated in Figure 2, the system is organized as a multi-stage pipeline with two parallel input streams. The structured stream combines historical traffic patterns and real-time traffic APIs into a unified representation; the unstructured stream extracts traffic-relevant events from web and social media content through a dedicated crawling and event-extraction module. The prediction module fuses the two streams and outputs travel time estimates for the road network, which are then consumed by the routing module to compute driving directions. In an operational setting the pipeline is iterative: the travel time observations collected while executing a route update the real-time traffic stream and, indirectly, the prediction module, so that subsequent predictions are refined on the fly.

Once accurate travel times have been estimated, the fundamental task of point-to-point routing can be formulated as a *Time-Dependent Quickest Path Problem* (TD-QPP) [18]. Formally, this problem is defined on a directed road network graph $G = (V, A)$, where V represents the set of nodes (e.g., intersections or junctions) and A the set of arcs (road segments connecting nodes). Each arc $(i, j) \in A$ is associated with a time-dependent travel time function $\tau_{ij}(t)$, which specifies the time required to traverse the arc when departing at time t. The goal of the TD-QPP is to determine the optimal path from a given origin to a destination that minimizes total travel time, taking into account the temporal variations encoded in the $\tau_{ij}(t)$ functions. Given a start time t and a route $p = (s = v_1, v_2, \ldots, v_k = d)$ from an origin node $s \in V$ to a destination node $d \in V$, its *traversal time* $z_p(t)$ is defined recursively as

$$z_{(v_1,\ldots,v_i)}(t) = z_{(v_1,\ldots,v_{i-1})}(t) + \tau_{(v_{i-1},v_i)}(t + z_{(v_1,\ldots,v_{i-1})}(t)), \tag{1}$$

with the initialization $z_{(v_1,v_2)}(t) = \tau_{(v_1,v_2)}(t)$. If the objective is to optimize path duration, the TD-QPP aims to find a path $p = (s = v_1, v_2, \ldots, v_k = d)$ that minimizes $z_p(t)$ as defined in Equation (1).

The TD-QPP is solvable in polynomial time on FIFO (First-In-First-Out, i.e., departing earlier results in arriving earlier) networks [19], but becomes NP-hard on non-FIFO

networks [20]. While time-invariant shortest path algorithms can be adapted for TD-QPP in principle, they are impractical in web applications that demand real-time responses on large-scale networks. This has led to specialized algorithms that incorporate road network features such as street hierarchies. For a comprehensive overview up to 2009, see [21]. A bidirectional A^* search using lower-bound arc costs to limit the search space was proposed in [22], achieving significant speed-ups over Dijkstra's algorithm. A different approach based on highway hierarchies for efficient, exact path computation in continental-scale networks was developed in [23]. Calogiuri, Ghiani and Guerriero, in their paper [24], demonstrated that under specific conditions, the TD-QPP can be reduced to a time-invariant problem using appropriately defined constant travel times. If these conditions do not hold, their approach yields a heuristic with a bounded approximation ratio. They also integrated an accurate lower bound [25] into a unidirectional A^* search, achieving reduced computation times on large metropolitan graphs. In public transportation networks, timetables define services as time-expanded graphs. For very large public transportation networks, Ref. [26] proposed effective models and developed algorithms capable of producing routes on graphs with up to 500 million arcs, yielding millisecond-scale query times. A shortest path problem in non-FIFO networks was addressed in [27] where the authors introduced a modified Dijkstra's algorithm that handles discontinuities effectively. On the other hand, Ref. [28] tackled shortest paths in multi-modal time-dependent networks under regular language constraints, proposing the *State-Dependent ALT* algorithm, which outperformed existing methods, while Ref. [29] developed a dynamic programming-based algorithm for solving the multi-criteria shortest path problem with time windows in multimodal, scheduled networks. Their method decomposes the problem into sub-problems solved using backward labelling, showing good performance on real-world scenarios. Heuristic adaptations of Dijkstra's algorithm for cost-minimizing paths in networks with time-dependent congestion and tolls were proposed in [30]. Their results showed significant cost sensitivity to traffic conditions and congestion pricing schemes.

3.1. Time-Dependent Least Consumption Path Problem

In the paper [31], the authors considered a setting where vehicle speed and fuel consumption vary based on peak traffic conditions. Each customer node has a time window and a fixed service time, with waiting allowed upon early arrival. Travel times are modeled using a two-phase speed function: a congested phase followed by a free-flow period. The goal is to minimize the total fuel and wage costs. They showed that for a fixed customer sequence, the problem is solvable in polynomial time.

3.2. Stochastic Time-Dependent Quickest Path Problem

Routing in stochastic time-varying networks can be approached in two ways: as a *fixed path* or as a *dynamic routing policy*. A fixed path, also known as static routing, is a pre-defined sequence of arcs, to be followed regardless of traffic realizations. This policy is generally suitable for small networks which rarely changes. A dynamic routing policy, in contrast, allows real-time decisions based on observed conditions, specifying the next arc to take based on current time and network state. This policy is used for large and very complex networks, characterized by frequent changes.

Under stochastic conditions, optimal routes are better represented as adaptive policies rather than fixed paths [32]. A dynamic programming approach was proposed in [33] for acyclic networks with exponential arc travel times and environmental conditions modelled by continuous-time Markov processes. Vehicles can wait for better environmental states before proceeding. The least expected time path problem, assuming independent, stochastic travel times, was addressed in [34,35]. They proposed a label-correcting algorithm with

exponential worst-case complexity, but near-linear performance in sparse networks. Time-adaptive policies were further examined in [36] where the authors proposed a heuristic algorithm that is guaranteed to be optimal when the heuristic provides an underestimate of the cost-to-go. The adaptive routing problem was investigated also in [37], where the authors proposed an algorithm to solve the problem in signalized stochastic, time-varying networks, where the signal timing plan and actual timings are known only probabilistically.

Optimal routing policy problems in stochastic time-dependent networks were analyzed in [38], proposing exact and approximate algorithms with a focus on computational efficiency. Finally, Ref. [39] developed a method for the a priori shortest path problem, which can be extended to compute the K shortest paths. Their algorithm demonstrated robustness and efficiency under realistic traffic distributions.

3.3. Centralized Driving Directions to Minimize Infrastructure Congestion and Risk

An alternative line of research addresses the challenge of centralized, intelligent traffic planning by dynamically diverting vehicles in real-time to reduce both localized risks and overall network congestion. The goal is to develop traffic management systems in which a central planner optimizes vehicle routes to protect vulnerable infrastructure, such as bridges, while simultaneously minimizing congestion across the monitored road network.

Major contributions in this field include [40] that developed a real-time optimization model that assigns vehicle paths to minimize risk to sensitive road structures. If vehicles are diverted, the algorithm seeks paths that lessen congestion elsewhere in the network. Computational experiments with real data confirmed the model's efficacy in preventing dangerous situations while managing overall congestion. A comprehensive review of methods bridging user equilibrium and system optimum in static traffic assignment was provided in [41], detailing approaches where a central coordinator assigns optimal paths to minimize road congestion. The paper offers algorithms and models relevant to centralized re-routing for congestion management, aligning with the need to balance individual and system-level routing objectives.

4. Travel Time Prediction Based on Structured Data

In this section, a review of the major commercial and governmental traffic data APIs is first proposed, highlighting the types of information they provide and their relevance for real-time traffic analysis. Then, a survey of the main travel time prediction techniques is presented, discussing the underlying methodologies, data requirements, and applicability to dynamic routing scenarios.

4.1. Structured Traffic Data APIs

Commercial and governmental traffic data APIs form the backbone of most real-time navigation systems. These services typically provide structured information such as current traffic speeds, congestion levels, incident reports, and estimated travel times. Data are generally collected from probe vehicles, road sensors, and crowd-sourced user reports, ensuring relatively high coverage and temporal resolution. Some of the most commonly used APIs are provided by Google Maps. For example, Ref. [42] used Google Maps APIs to collect real-world, recent speed data across defined geographic areas, enabling the study of urban traffic behavior and speed estimation based on both time of day and day of week. Their study, validated in Boston (MA, USA), identified congestion clusters and provided managerial insights for improving network efficiency. Similarly, Ref. [43] employed the Google Maps Distance Matrix API for real-time incident detection by continuously analyzing traffic data and identifying anomalies through a mathematical model that compared current, regular, and historical traffic flows. In [44], the authors used travel time data from

the Google Maps API to study temporal and spatial congestion patterns in Jeddah, Saudi Arabia (a city with narrow, high-traffic corridors and limited public transit). They defined a real-time congestion index to quantify traffic flow.

Another widely adopted platform is Waze, which relies on crowd-sourced reports from its large community of users. The Waze API was used in [45] for a spatio-temporal analysis of congestion patterns in Casablanca (Morocco), applying clustering algorithms to classify areas into congestion categories, and a Naïve Bayes classifier to examine land-use factors (e.g., residential density, industrial areas, and tramways) influencing congestion. Waze traffic speed data were used in [5] and compared with remote traffic microwave sensors, showing that Waze speeds were more reliable during congested conditions due to higher observation density. The reliability and coverage of Waze data against the advanced traffic management system in Iowa (USA) were evaluated in [46], demonstrating that Waze provides broad and timely incident coverage, making it a valuable complement to traditional monitoring systems. Crowdsourced Waze user reports were further analyzed in [47] for secondary crash detection on I-40 in Knoxville, showing that Waze reports could distinguish primary and secondary crashes, some of which were missed by traditional speed contour methods.

Beyond APIs, IoT sensors represent another essential data source for urban mobility. Ref. [48] emphasized their transformative role in transport management, as sensors embedded in traffic, environmental, vehicular, and infrastructure systems provide real-time information on traffic patterns, air quality, road conditions, and parking availability. These data support accident prediction, congestion monitoring, and optimization of traffic signals. Within the PRONTO project, Ref. [49] demonstrated that data from on-board vehicle sensors (GPS, noise, temperature, acceleration, CAN bus/FMS interface) can be analyzed to detect anomalies and threats in vehicles, supporting data fusion, machine learning, and event recognition in public transport contexts.

Travel Time Modelling

In this section we briefly describe how arc travel time functions $\tau_{ij}(t)$ can be modelled. A dominant approach in the literature is the Ichoua, Gendreau, and Potvin (IGP) model [50]. The planning horizon is divided into discrete time intervals $[T_h, T_{h+1}]$, and the speed on arc (i, j) during interval h is assumed constant and denoted v_{ijh}. Here, travel time functions $\tau_{ij}(t)$ are computed by simulating vehicle movement through each interval with an iterative procedure that updates the remaining distance, time, and speed slot until the arc is completely traversed. As a result, the IGP travel time functions are continuous and piecewise linear. It has been shown in [51] that any continuous piecewise linear function that satisfies the FIFO property can be accurately represented within this model. Moreover, the parameters of the IGP model can be efficiently computed by solving a set of linear systems, making it both expressive and computationally tractable for time-dependent routing applications.

4.2. Speed-Based Prediction Pipeline

Several prediction methods exploit the variability exhibited by traffic conditions due to periodic (e.g., hourly, weekly) and random (e.g., weather, accidents) effects [52]. The process is typically made up of three phases: (1) predicting speeds; (2) computing arc travel times; (3) querying quickest paths (i.e., estimating travel times between given origin–destination pairs).

4.2.1. Speed Prediction

Speed prediction relies on long-term GPS data, which include timestamps, vehicle IDs, locations, and speeds. The process consists of three main stages:

- *Data Preparation*: GPS points are mapped to road arcs using map-matching algorithms [53]. Abnormal or parking-related points are filtered out. A complete daily speed profile can require over 100 million speed patterns [54].
- *Size Reduction and Clustering*: Inadequate patterns (e.g., less than 5% data coverage) are removed. Speed patterns are aggregated (e.g., 15-min intervals grouped into one-hour bins). Clustering techniques, such as K-means followed by Affinity Propagation [11,12], are used to classify arc behavior.
- *Prediction*: Methods are categorized into:
  - *Naive*: Simple averages, low accuracy.
  - *Parametric*: Use models like ARIMA, Kalman Filters.
  - *Non-parametric*: Data-driven models like SVMs, ANNs, RNNs, and LSTMs.

Table 1 provides a comparative overview of the main categories of prediction methods discussed above, summarizing their data requirements, typical accuracy, computational complexity, and key limitations. Note that, for each time prediction approach, a qualitative assessment (low, moderate, good, high) of accuracy and complexity is furnished.

Table 1. Comparative overview of travel time prediction approaches.

Approach	Data Requirements	Accuracy	Complexity	Key Limitations
Naive (averages)	Historical speed profiles	Low	Very low	Cannot capture temporal dynamics
ARIMA	Time-series speed/flow data	Moderate	Low	Assumes stationarity; limited for non-linear patterns
Kalman Filters	Sequential observations with state model	Moderate	Low–moderate	Sensitive to model assumptions
SVM	Feature-engineered traffic data	Moderate–good	Moderate	Requires careful feature selection; limited scalability
ANN	Historical traffic features	Good	Moderate	Prone to overfitting on small datasets
RNN/LSTM	Sequential time-series data	Good–high	High	Training time; vanishing gradient in long sequences
GNN	Graph-structured road network + traffic data	High	High	Requires explicit graph construction; computationally intensive
Transformer	Large-scale spatiotemporal data	High	Very high	Data-hungry; high computational cost

4.2.2. Computing Travel Times

Two primary query types are relevant:

- *Travel Time Queries*: Compute $\Phi_{ij}(t)$, arrival time at node j from arc (i,j) at departure time t, using

$$\Phi_{ij}(t) = \left\{ x \,\middle|\, \int_t^x v_{ij}(t')dt' = d_{ij} \right\}, \quad (2)$$

  where d_{ij} is the length of arc (i,j). Vidal et al. [55] proposed storing $\Phi_{ij}(t)$ in Equation (2) as closed-form piecewise-linear (PL) functions, allowing $O(1)$ or $O(\log H)$ access.
- *Quickest Path Queries*: Compute earliest arrival $\Gamma_{ij}(t)$ from i to j, considering time-dependent paths. Vidal et al. (2021) [55] avoid the inefficiency of repeated shortest path queries via preprocessed closed-form $\Gamma_{ij}(t)$ functions.

Despite the theoretical exponential complexity [56], practical instances exhibit few breakpoint changes, making storage and computation feasible.

4.2.3. Approximating Travel Times for Routing

Approximate travel times can accelerate optimization algorithms. For instance, in [57] the authors estimate delay penalties for local route modifications. A notable method is based on the *speed decomposition* by [58], which factors speeds as follows:

$$v_{ijh} = u^0_{ij} b^0_h \delta^0_{ijh}, \tag{3}$$

where u^0_{ij} is the arc's max speed, b^0_h is the time-dependent congestion factor, and δ^0_{ijh} is a degradation factor. In Equation (3), if all arcs have $\delta^0_{ijh} = 1$, the network exhibits *path ranking invariance*, i.e., path order is constant over time. This approximation is improved in [59] by selecting greater u_{ij} values and recalculating $b_h(\mathbf{u})$ and $\delta_{ijh}(\mathbf{u})$, leading to a new lower bound:

$$v_{ijh} = u_{ij} b_h(\mathbf{u}) \delta_{ijh}(\mathbf{u}), \tag{4}$$

where $\mathbf{u}$ is a vector of u_{ij}.

In Equation (4), the trade-off is between decreasing duration and increasing congestion penalties. A further refinement of this approach is proposed in [60] by selecting time intervals where the approximation $\underline{\tau}_{ij}(t)$ equals the true $\tau_{ij}(t)$ at least once, yielding tighter lower bounds.

5. Travel Time Prediction Based on Web and Social Media Events

In this section, we begin by reviewing techniques used for crawling the web and social media to automatically detect events that can impact traffic conditions (e.g., accidents, road closures, public gatherings, protests, or severe weather). We highlight the types of information these sources provide, such as location, time, severity, and event category, and discuss their relevance for real-time traffic analysis and incident-aware routing. We then present a survey of the main travel time prediction techniques that exploit such event-based data, examining their underlying methodologies, data requirements, and effectiveness in capturing the dynamic and often unpredictable nature of urban traffic. Particular attention is given to the integration of unstructured textual information with traditional traffic data, and the challenges associated with event extraction, temporal reasoning, and spatial alignment.

5.1. Semantic Crawling of Web and Social Media for Event Identification and Classification

To complement structured APIs and IoT sensors, semantic crawling techniques [61] are increasingly employed to gather unstructured but context-rich information from online sources such as news outlets, traffic authority websites, and social media platforms (e.g., X/Twitter, Facebook, or online forums). These sources often provide early warnings of accidents, protests, weather-related disruptions, or special events that may not yet appear in structured API data. Effective crawlers must implement filtering and prioritization strategies to handle the high volume and heterogeneity of textual streams, ensuring that only relevant and geo-specific information is retained.

Several studies demonstrate the value of social media for traffic event detection. A cost-effective method for incident detection by mining tweets was proposed in [62], filtering them using keyword-based features, classifying them as incident-related or not, geocoding locations, and categorizing incidents into five types. A framework for real-time incident detection using Twitter data was introduced in [63], where sentiment and emotion analysis were adopted to capture user perceptions. The works in [64,65] focused on Saudi-Arabic tweets, presenting machine learning and big-data frameworks that detect accidents, closures, weather disruptions, and public celebrations with high accuracy. The same idea was extended in [66] to Spanish-language tweets, extracting accident locations in Bogotá (Colombia) via Named Entity Recognition (NER) and geocoding. A social network-based

framework for traffic accident detection and condition analysis from Twitter and Facebook was developed in [4], extracting structured traffic information from unstructured posts. Moreover, an automated framework was designed in [67], using NLP for traffic monitoring that mines social media data to detect relevant events and issue timely alerts. A fine-tuned BERT (Bidirectional Encoder Representations from Transformers) model filtered and classified the data, while a Question-Answering module extracted key details such as location, time, and event type.

Other works highlight large-scale applications. In particular, the authors in [68] processed 120,000 geo-tagged traffic tweets and six million non-traffic tweets from the U.S., building classification models and performing spatio-temporal analyses that showed social media reflects real traffic patterns. Also, in [69] a huge amount of user-generated Twitter data was investigated and a novel latent Dirichlet allocation (LDA) approach ("tweet-LDA") was proposed for instant traffic alert generation and geocoding. On the other hand, the authors in [70] used text mining and NER on Indonesian tweets to monitor road conditions, evaluating classifiers with Bag-of-Words and TF-IDF features, while Ref. [71] presented an NLP approach based on Support Vector Machines for incident detection in the UK, and Ref. [72] combined deep learning and sentiment-weighted Kernel Density Estimation on Chinese microblogs (Sina Weibo) to map accident- and congestion-prone areas. Deep learning techniques were also implemented in [73] to develop a novel traffic management approach. The model was specifically applied to Twitter data to notify relevant authorities about critical issues such as road accidents and traffic congestion within the Chandigarh tri-city region.

While the studies above reviewed predominantly rely on Twitter/X as the primary social media source, the digital landscape for real-time event reporting has evolved considerably. Platforms such as TikTok, Telegram, and Reddit are increasingly used to share information about traffic disruptions. TikTok presents unique challenges due to its video-first format, which requires computer vision techniques rather than traditional NLP for event extraction. Telegram hosts dedicated channels for traffic alerts in several cities, offering relatively structured and timely information [74]. Reddit, through location-specific subreddits, provides community-sourced accounts of local traffic conditions. Similarly, Facebook has been exploited for accident event detection in non-English contexts [75]. Overall, the maturity of transportation-analytics use differs significantly across these platforms: Twitter/X and Waze have a consolidated body of published applications, Facebook and Telegram have been used in specific and relatively recent case studies [74,75], while TikTok and Reddit remain largely at the stage of emerging potential, with few systematic transportation studies available to date. However, restricted API access and the heterogeneity of content formats across these platforms remain significant barriers to their systematic integration into traffic monitoring systems.

The use of social media data for traffic monitoring also raises important ethical concerns [76,77]. Mining user-generated content involves privacy risks, as individuals may not be aware that their posts are being used for traffic surveillance purposes [78]. Algorithmic bias may emerge in areas with lower social media penetration, leading to uneven monitoring coverage across socio-economic groups and neighborhoods [79]. Furthermore, compliance with data protection regulations such as the European General Data Protection Regulation (GDPR) imposes constraints on the collection, storage, and processing of personal data in transportation contexts [80]. Addressing these concerns is necessary before social media-based traffic monitoring can be deployed at scale.

One aspect of this literature that is rarely quantified is end-to-end detection latency. Many of the frameworks cited above claim to be "real-time" systems and rely on streaming infrastructures such as the Twitter Streaming API (v1.1) or Apache Spark (v2.4–v3.x)

pipelines. Their evaluations, however, focus on offline metrics (precision, recall, F1-score, accuracy) computed on historical datasets, with little reporting of the time elapsed between an event occurring and its detection by the system. In navigation applications, routing decisions must be made within seconds; without comparable latency benchmarks, it is difficult to assess which of these methods actually meet the real-time requirements they claim. We list this gap among the open research directions in Section 6.

5.2. Event-Based Travel Time Prediction

Recent advances in travel time prediction employ various integration strategies to merge event-based data from web and social media with traditional traffic datasets. These methods aim at capturing the dynamic disruptions observed in urban environments caused by accidents, protests, and weather events. In this section, we examine four prominent categories of integration, graph neural network architectures, attention mechanisms, sequence-to-sequence temporal models, and hybrid frameworks, with emphasis on their methodologies, data fusion approaches, and effectiveness.

For a broader methodological overview of spatio-temporal traffic prediction, see the survey by Yuan and Li (2021) [16].

5.2.1. Graph Neural Network Architectures

Graph Neural Networks (GNNs) [81] have the potential to represent the complexity of traffic dynamics by mapping event reports (extracted from unstructured web or social media text) directly onto nodes and edges of a graph via geotagging and NLP-based location inference. GNNs allow multi-modal data fusion from traffic sensors, weather feeds, and textual event streams by encoding each heterogeneous source as a distinct class or type of node/edge. Event impacts can be rapidly propagated across the network, enabling prediction models to capture both localized and distributed disruptions. The spatial context means that a single incident can affect connected segments, with concurrency priors and network-wide inference enhancing predictive accuracy. See [82] for a recent study that proposes a spatio-temporal fusion graph convolutional network for traffic prediction, incorporating weather factors. A further example is the physics-guided stepwise GNN framework of Pan et al. (2025) [17], which targets traffic flow prediction at urban intersections.

Taken together, GNN-based approaches are well suited to settings where the road network topology is known and multi-modal features (sensor readings, weather, event reports) can be attached to nodes or edges. Their main strengths are the explicit modelling of spatial dependencies and the ability to propagate event impacts across connected segments. The main limitations are the computational cost on very large networks, the need for substantial labelled data, and the difficulty of interpreting the effect of specific inputs on predictions.

5.2.2. Attention Mechanisms in Spatiotemporal Fusion

Attention mechanisms embedded in deep learning architectures enable selective focus on events, road segments, and time windows most influential for traffic prediction. Multi-modal attention networks dynamically weigh the importance of incoming clusters of events, traffic sensors, and weather patterns in both the spatial and temporal domains, amplifying the effect of incidents with outsized impact. Sparse and hierarchical attention allows these models to prioritize data sources, isolating network bottlenecks during critical disruptions, which is vital for context-aware forecasting. See [83,84] for two recent contributions in this field.

Attention-based approaches are attractive when the relative importance of events, road segments, and time windows changes across contexts, because they adjust their input

weighting dynamically. A secondary benefit is a degree of intrinsic interpretability through the inspection of attention weights. The main drawbacks are sensitivity to hyperparameter tuning, a tendency to overfit on sparse event streams, and a computational cost that scales super-linearly with sequence length.

5.2.3. Sequence-to-Sequence Temporal Reasoning

Sequence-to-sequence (S2S) models process detected event sequences from web and social media, which often have incomplete or noisy temporal information. Using components like multi-head and additive (Bahdanau) attention, these models infer the temporal dynamics of event impacts (start time, duration, cascading effects) and predict corresponding traffic fluctuations. The integration of NLP-driven time inference allows S2S models to convert fragmented reports into coherent temporal disruptions, enabling temporally aligned traffic predictions that reflect real-world evolutionary patterns (see, e.g., [85]).

Sequence-to-sequence models fit naturally when the temporal evolution of an event (onset, peak, dissipation) is the primary predictive signal and inputs have variable length. They are well suited to reasoning about cascading effects across time, but are harder to parallelize than attention-only models, can struggle with very long temporal horizons, and depend heavily on accurate alignment between event reports and sensor signals.

5.2.4. Hybrid Fusion Frameworks

Hybrid fusion frameworks combine GNN, attention, and sequence-based reasoning in unified end-to-end systems. These architectures integrate structured sensor data, unstructured text (event reports, web news, social feeds), geospatial metadata, and historical traffic records. Multi-source fusion Graph Convolutional Networks, for example, process sensor readings alongside event-extracted spatial features and environmental conditions. These models tend to be more tolerant of noisy inputs and better able to adapt to sudden disruptions, outperforming single-source or traditional time-series models, particularly in event-rich environments (see [86] for a recent contribution in this area).

Recent contributions have further advanced spatiotemporal traffic forecasting through specialized deep learning architectures. Cai et al. [13] proposed JointSTNet, a joint pre-training framework that employs graph capsule modules to capture complex spatial dependencies across urban road networks. Zhang et al. [14] introduced a two-way heterogeneity model based on dynamic graph convolution, which explicitly handles anomalous events such as traffic accidents by modelling both spatial and temporal heterogeneity. Lin et al. [15] developed GSA-KAN, a hybrid model that combines Kolmogorov–Arnold Networks with a Gravitational Search Algorithm for short-term traffic forecasting, demonstrating improved accuracy over conventional approaches.

Overall, hybrid frameworks tend to perform best in event-rich environments with heterogeneous data sources, because they can exploit the complementary strengths of graph structure, attention, and sequential reasoning. The trade-off is higher architectural complexity, longer training times, and more hyperparameters to tune. Overfitting becomes a practical concern when the event data are sparse or imbalanced across the network.

Table 2 provides a classification of the event-based prediction studies reviewed in this section, organized by integration approach, forecasting target, data sources, and key contribution.

A related concern is the interpretability of these event-based models. Attention-based approaches, including those by [83–85] offer a first form of intrinsic interpretability: their attention weight matrices can be inspected to reveal which time steps, sensors, or event types drive a given prediction. The authors in [15] further argue that their KAN-based architecture, with spline-parametrized univariate functions along the network edges, is

more interpretable than a standard multilayer perceptron. Post-hoc techniques such as LIME, SHAP, and counterfactual analysis for traffic prediction are discussed in the recent survey by [8]. Earlier, Ref. [7] framed deep neural traffic models as "black-box" and identified interpretability as an open gap. The systematic use of explainable AI in event-aware travel time prediction, however, remains limited; we include it in the future research directions of Section 6.

Table 2. Classification of event-based travel time prediction approaches.

Category	Reference	Forecasting Target	Data Sources	Key Contribution
GNN	[81]	GNN survey (methodological)	Multi-modal (sensors, text, weather)	Comprehensive GNN survey
GNN	[82]	Weather-aware traffic flow	Sensors + weather	Spatio-temporal fusion GCN
GNN	[14]	Incident-aware traffic flow	Sensors + events	Two-way heterogeneity with dynamic graph conv.
Attention	[83]	Long/short-term traffic flow	Multi-modal traffic data	Multi-modal attention for long-short term correlation
Attention	[84]	Spatial-temporal traffic flow	Traffic sensors	Hierarchical attention mechanism
S2S	[85]	Multistep traffic speed	Time-series + events	Multi-step temporal reasoning
Hybrid	[86]	Urban traffic flow (multi-source)	Sensors + text + spatial	Multi-source fusion GCN
Hybrid	[13]	Spatio-temporal traffic forecasting	Spatiotemporal traffic	Joint pre-training with graph capsules
Hybrid	[15]	Short-term traffic flow	Short-term traffic data	KAN + gravitational search algorithm

Remarks

Traditional routing algorithms face significant scalability limitations when applied to contemporary large-scale urban networks. While time-invariant shortest path algorithms can theoretically be adapted for the TD-QPP, they are impractical in web applications that demand real-time responses on large-scale networks. Moreover, whereas the TD-QPP is solvable in polynomial time on FIFO networks, it becomes NP-hard on non-FIFO networks, making traditional methods difficult to scale when network conditions violate the FIFO property. Scalability is further hampered by the sheer volume of data required to achieve accurate modelling. For instance, generating a complete daily speed profile for an urban network can require processing over 100 million speed patterns, necessitating significant data preparation, filtering, and clustering to remain computationally tractable. To address these limitations, several specialized techniques have been developed:

- **Heuristics and Hierarchies:** Algorithms using highway hierarchies or bidirectional searches with lower-bound arc costs achieve significant speed-ups over standard approaches;
- **Pre-processed Functions:** Storing travel times as pre-processed closed-form piecewise-linear functions allows for $O(1)$ or $O(\log H)$ access times, avoiding the inefficiency of repeated shortest path queries;
- **Public Transit Optimization:** For extremely large networks, such as public transportation systems with up to 500 million arcs, the use of transfer patterns and time-expanded graphs enables millisecond-scale query times.

The scaling of complex deep learning models for online deployment remains a primary challenge. Moving beyond simple local delay warnings toward robust, city-wide adaptive systems will require further progress in data extraction pipelines and interpretable learning

architectures. In this context, centralized traffic management, where a central planner optimizes routes to minimize overall network congestion rather than just individual travel times, is gaining increasing attention as a viable approach to urban-scale routing.

6. Conclusions and Future Research Directions

This survey has examined recent developments in real-time navigation systems, tracing the shift from traditional methods based on structured data to advanced deep learning approaches that integrate both structured traffic data and unstructured contextual information from sources like news and social media. Only recently have some navigation systems and apps begun incorporating local events—such as strikes, concerts, sports games, and political rallies—into their ETA predictions. However, these systems typically go no further than issuing delay warnings near venues, like stadiums during major events. This paper reviews key methodologies in the field, including both well-established approaches and recent innovations. It covers techniques ranging from natural language processing for event extraction, to machine learning for traffic forecasting, and graph-based algorithms for efficient routing, highlighting their respective strengths and limitations.

By jointly covering time-dependent routing, structured and unstructured traffic data, and NLP-based event extraction within a unified framework, this survey connects three research communities that have historically worked largely independently of one another. The end-to-end pipeline perspective adopted throughout the paper highlights the interdependencies among data acquisition, predictive modelling, and route optimization, offering a reference point for researchers across operations research, machine learning, and natural language processing.

The surveyed literature also grounds these methodologies in concrete real-world applications. Structured-API studies have been applied to Boston, MA [42], and Jeddah, Saudi Arabia [44]; Waze-based analyses cover Casablanca, Morocco [45], Knoxville, TN [5,47], and Iowa [46]; Twitter-based event detection has been demonstrated on Arabic [64,65], Spanish [66], English-language [4,67], and Indian [73] corpora. This spread of settings, languages, and data sources shows that the techniques reviewed here are not only conceptual but already tested in operational conditions, although often in isolated case studies rather than in integrated navigation systems.

Policy Implications

The methodologies surveyed in this paper have direct implications for smart-city stakeholders, transport agencies, and urban planners. First, fusing structured traffic data with unstructured textual streams can support evidence-based decision making, for example by enabling data-driven prioritization of incident response and of information dissemination to commuters. Second, the use of social media data for traffic monitoring requires explicit regulatory frameworks that balance operational utility with privacy protection, notably under the GDPR in the European context. Third, context-aware routing algorithms can be used to implement congestion-pricing and low-emission zone policies that adapt to actual network-wide conditions rather than to static rules. Finally, the deployment of such systems calls for coordination between municipal traffic authorities, software providers, and civil-society stakeholders, so that the benefits of data-driven navigation translate into concrete improvements in urban mobility.

Several limitations still affect current traffic-aware navigation systems. Data latency remains a persistent issue, as API refresh rates and the inherent delays of social media reporting hinder real-time adaptability. Closely related is the challenge of event severity quantification: while many models can detect the occurrence of an incident, determining whether it significantly impacts travel times remains largely unresolved. Scalability poses a

major barrier as well, since semantic crawlers must operate across multiple languages and cities without degradation in accuracy. Furthermore, systems calibrated in a specific urban environment often fail to transfer effectively to other contexts with different infrastructures and mobility dynamics, limiting their generalization capacity [8].

Related to this, the event-detection literature rarely reports end-to-end detection latency as an evaluation metric, making it difficult to compare methods on the real-time performance they often claim.

Practical deployment challenges also deserve attention. The interoperability between legacy traffic management systems and modern data-driven platforms is often limited, creating integration bottlenecks in real-world scenarios. The computational and energy costs associated with large-scale real-time processing raise sustainability concerns that are frequently overlooked in the smart mobility literature. Additionally, ethical considerations surrounding the mining of social media data for traffic monitoring, including privacy risks, algorithmic bias in areas with lower social media penetration, and compliance with regulations such as the GDPR [77,80], must be addressed in the design of future systems.

We group the main future research directions into three categories: (i) *methodological*, including multimodal fusion frameworks, cross-city transferable models, and explainable AI for traffic prediction; (ii) *computational*, including scalable online deployment of deep learning models, geolocating ambiguous event mentions, and energy-aware processing; and (iii) *deployment-oriented*, including interoperability with legacy traffic management systems, user feedback loops, privacy-preserving mechanisms, and integration with smart mobility ecosystems. In the following paragraphs we elaborate on each of these strands.

Future research should focus on geolocating ambiguous event mentions, resolving temporal uncertainty in social data, and scaling complex models for online deployment. To this aim, gazetteer-based methods and services such as OpenStreetMap Nominatim provide broad geographic coverage, while machine learning-based geoparsing frameworks [87–89] offer improved disambiguation capabilities for informal text. Error-handling mechanisms, including confidence scoring and spatial validation against known road networks, are essential for ensuring the reliability of geocoded event locations extracted from noisy social media streams. In addition, further developments should prioritize the design of scalable multimodal fusion frameworks capable of balancing accuracy and computational efficiency [7,9], together with the development of cross-city transferable models that generalize across heterogeneous urban infrastructures. The integration of user feedback loops to continuously refine predictions and routing decisions, the embedding of privacy-preserving mechanisms into data pipelines, and the advancement of explainable AI methods to increase trust in system outputs should also be explored. In particular, attention-weight visualization, SHAP-based feature attribution, and counterfactual explanations tailored to event-aware traffic models deserve systematic investigation. On a broader scale, real-time reinforcement learning and the integration of navigation with smart mobility ecosystems, including electric vehicle charging optimization and multimodal transport planning, are yet to be fully addressed.

Author Contributions: Conceptualization, G.G.; resources and data curation, G.R., G.G., E.M. and V.M.; writing—original draft preparation, G.G., E.M., V.M., S.D.I., M.P. and G.R.; writing—review and editing, G.G., E.M., V.M., S.D.I., M.P. and G.R.; supervision, G.G., S.D.I. and M.P.; funding acquisition, G.G., S.D.I. and M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the following research programs: (a) "Sustainable Mobility Center" (Centro Nazionale per la Mobilità Sostenibile—CN MOST), Project Code CN00000023, Spoke 7; (b) National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.4—Call for tender No. 3138 of 16 December 2021, rectified by Decree n.3175 of 18 December 2021 of Italian Ministry of University and Research funded by the European Union-Next GenerationEU.

Award Number: Project Code CN00000033, Concession Decree No. 1034 of 17 June 2022 adopted by the Italian Ministry of University and Research, CUP F87G22000290001, Project title "National Biodiversity Future Center—NBFC"; (c) ICSC-National Research Center in High Performance Computing, Big Data and Quantum Computing, funded by the European Union-NextGenerationEU, Project name: PNRR-HPC; Project Code: CN00000013; CUP: C83C22000560007.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT (GPT-5, OpenAI; https://chat.openai.com) exclusively for language polishing, specifically grammar correction and stylistic refinements of the English text, applied uniformly across all sections of the manuscript. ChatGPT was not used for literature search, content generation, data analysis, or the formulation of any finding or conclusion presented in this work. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

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

Appendix A. Full Boolean Search Expressions

This appendix reports the full Boolean expressions adopted for the systematic search in Scopus. The queries were adapted to the syntax of Web of Science and IEEE Xplore while preserving the same terms and logical operators. All queries were restricted to the 2000–2025 publication window through the filter PUBYEAR $>$ 1999 AND PUBYEAR $<$ 2026.

- *Query A (time-dependent routing):* TITLE-ABS-KEY(("time-dependent" OR "time-varying") AND ("shortest path" OR "vehicle routing" OR "quickest path" OR "fastest path" OR "travel time" OR "arc routing" OR "route planning" OR "road network")).
- *Query B (traffic prediction):* TITLE-ABS-KEY(("traffic flow prediction" OR "traffic forecasting" OR "traffic speed prediction" OR "travel time prediction" OR "traffic congestion prediction") AND ("deep learning" OR "neural network" OR "graph neural" OR "machine learning" OR "LSTM" OR "attention" OR "spatio-temporal" OR "spatiotemporal")).
- *Query C1 (social media and traffic):* TITLE-ABS-KEY(("social media" OR "Twitter" OR "Waze" OR "crowdsourc*") AND ("traffic incident" OR "traffic accident" OR "traffic event" OR "traffic congestion" OR "road incident" OR "road accident")).
- *Query C2 (NLP and traffic):* TITLE-ABS-KEY(("NLP" OR "natural language processing" OR "text mining" OR "geopars*" OR "toponym") AND ("road traffic" OR "traffic accident" OR "traffic incident" OR "traffic congestion" OR "traffic event" OR "traffic monitoring" OR "urban traffic")).

References

1. International Transport Forum. *Monitoring Progress in Urban Road Safety: 2022 Update*; Technical report; OECD Publishing: Paris, France, 2022. [CrossRef]
2. Yang, M.; Liu, Y.; You, Z. The Reliability of Travel Time Forecasting. *IEEE Trans. Intell. Transp. Syst.* **2010**, *11*, 162–171. [CrossRef]
3. Chen, C.Y.T.; Sun, E.W.; Chang, M.F.; Lin, Y.B. Enhancing travel time prediction with deep learning on chronological and retrospective time order information of big traffic data. *Ann. Oper. Res.* **2024**, *343*, 1095–1128. [CrossRef] [PubMed]
4. Ali, F.; Ali, A.; Imran, M.; Naqvi, R.A.; Siddiqi, M.H.; Kwak, K.S. Traffic accident detection and condition analysis based on social networking data. *Accid. Anal. Prev.* **2021**, *151*, 105973. [CrossRef]
5. Zhang, Z.; Han, L.D.; Liu, Y. Exploration and evaluation of crowdsourced probe-based Waze traffic speed. *Transp. Lett.* **2022**, *14*, 546–554. [CrossRef]

6. Wang, S.; Zhang, Y.; Hu, Y.; Yin, B. Knowledge fusion enhanced graph neural network for traffic flow prediction. *Phys. A Stat. Mech. Its Appl.* **2023**, *623*, 128842. [CrossRef]
7. Tedjopurnomo, D.A.; Bao, Z.; Zheng, B.; Choudhury, F.M.; Qin, K. A survey on modern deep neural network for traffic prediction: Trends, methods and challenges. *IEEE Trans. Knowl. Data Eng.* **2022**, *34*, 1544–1561. [CrossRef]
8. Du, K.; Guo, X.; Li, L.; Song, J.; Shi, Q.; Hu, M.; Fang, J. Traffic prediction in time series, spatial-temporal, and OD data: A systematic survey. *J. Traffic Transp. Eng. Engl. Ed.* **2025**, *12*, 666–700. [CrossRef]
9. Jiang, W.; Luo, J. Graph neural network for traffic forecasting: A survey. *Expert Syst. Appl.* **2022**, *207*, 117921. [CrossRef]
10. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef] [PubMed]
11. MacQueen, J. Classification and analysis of multivariate observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Los Angeles, CA, USA, 21 June–18 July 1965 and 27 December 1965–7 January 1966*; University of California Press: Berkley, CA, USA, 1967; pp. 281–297.
12. Frey, B.J.; Dueck, D. Clustering by Passing Messages Between Data Points. *Science* **2007**, *315*, 972–976. [CrossRef]
13. Cai, D.; Chen, K.; Lin, Z.; Li, D.; Zhou, T.; Leung, M.F. JointSTNet: Joint Pre-Training for Spatial-Temporal Traffic Forecasting. *IEEE Trans. Consum. Electron.* **2025**, *71*, 6239–6252. [CrossRef]
14. Zhang, H.; Lin, Z.; Xie, H.; Zhou, J.; Song, Y.; Zhou, T. Two-way heterogeneity model for dynamic spatiotemporal traffic flow prediction. *Knowl.-Based Syst.* **2025**, *320*, 113635. [CrossRef]
15. Lin, Z.; Wang, D.; Cao, C.; Xie, H.; Zhou, T.; Cao, C. GSA-KAN: A Hybrid Model for Short-Term Traffic Forecasting. *Mathematics* **2025**, *13*, 1158. [CrossRef]
16. Yuan, H.; Li, G. A Survey of Traffic Prediction: From Spatio-Temporal Data to Intelligent Transportation. *Data Sci. Eng.* **2021**, *6*, 63–85. [CrossRef]
17. Pan, Y.A.; Li, F.; Li, A.; Niu, Z.; Liu, Z. Urban intersection traffic flow prediction: A physics-guided stepwise framework utilizing spatio-temporal graph neural network algorithms. *Multimodal Transp.* **2025**, *4*, 100207. [CrossRef]
18. Adamo, T.; Gendreau, M.; Ghiani, G.; Guerriero, E. A review of recent advances in time-dependent vehicle routing. *Eur. J. Oper. Res.* **2024**, *319*, 1–15. [CrossRef]
19. Kaufman, D.E.; Smith, R.L. Fastest paths in time-dependent networks for intelligent vehicle-highway systems application. *J. Intell. Transp. Syst.* **1993**, *1*, 1–11. [CrossRef]
20. Orda, A.; Rom, R. Minimum weight paths in time-dependent networks. *Networks* **1991**, *21*, 295–319. [CrossRef]
21. Delling, D.; Wagner, D. Time-Dependent Route Planning. In *Robust and Online Large-Scale Optimization: Models and Techniques for Transportation Systems*; Ahuja, R.K., Möhring, R.H., Zaroliagis, C.D., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 207–230. [CrossRef]
22. Nannicini, G.; Delling, D.; Schultes, D.; Liberti, L. Bidirectional A^* search on time-dependent road networks. *Networks* **2012**, *59*, 240–251. [CrossRef]
23. Delling, D. Time-Dependent SHARC-Routing. *Algorithmica* **2011**, *60*, 60–94. [CrossRef]
24. Calogiuri, T.; Ghiani, G.; Guerriero, E. The time-dependent quickest path problem: Properties and bounds. *Networks* **2015**, *66*, 112–117. [CrossRef]
25. Ghiani, G.; Guerriero, E. A lower bound for the quickest path problem. *Comput. Oper. Res.* **2014**, *50*, 154–160. [CrossRef]
26. Bast, H.; Carlsson, E.; Eigenwillig, A.; Geisberger, R.; Harrelson, C.; Raychev, V.; Viger, F. Fast Routing in Very Large Public Transportation Networks Using Transfer Patterns. In *Algorithms—ESA 2010*; de Berg, M., Meyer, U., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 290–301. [CrossRef]
27. Dell'Amico, M.; Iori, M.; Pretolani, D. Shortest paths in piecewise continuous time-dependent networks. *Oper. Res. Lett.* **2008**, *36*, 688–691. [CrossRef]
28. Barrett, C.; Jacob, R.; Marathe, M. Formal-Language-Constrained Path Problems. *SIAM J. Comput.* **2000**, *30*, 809–837. [CrossRef]
29. Androutsopoulos, K.N.; Zografos, K.G. Solving the multi-criteria time-dependent routing and scheduling problem in a multimodal fixed scheduled network. *Eur. J. Oper. Res.* **2009**, *192*, 18–28. [CrossRef]
30. Wen, L.; Catay, B.; Eglese, R. Finding a minimum cost path between a pair of nodes in a time-varying road network with a congestion charge. *Eur. J. Oper. Res.* **2014**, *236*, 915–923. [CrossRef]
31. Franceschetti, A.; Honhon, D.; Laporte, G.; Van Woensel, T. A Shortest-Path Algorithm for the Departure Time and Speed Optimization Problem. *Transp. Sci.* **2018**, *52*, 756–768. [CrossRef]
32. Hall, R.W. The Fastest Path through a Network with Random Time-Dependent Travel Times. *Transp. Sci.* **1986**, *20*, 182–188. [CrossRef]
33. Azaron, A.; Kianfa, F. Dynamic shortest path in stochastic dynamic networks: Ship routing problem. *Eur. J. Oper. Res.* **2003**, *144*, 138–156. [CrossRef]

34. Miller-Hooks, E.D.; Mahmassani, H.S. Least possible time paths in stochastic, time-varying networks. *Comput. Oper. Res.* **1998**, *25*, 1107–1125. [CrossRef]
35. Miller-Hooks, E.; Mahmassani, H. Path comparisons for a priori and time-adaptive decisions in stochastic, time-varying networks. *Eur. J. Oper. Res.* **2003**, *146*, 67–82. [CrossRef]
36. Bander, J.L.; White, C.C. A Heuristic Search Approach for a Nonstationary Stochastic Shortest Path Problem with Terminal Cost. *Transp. Sci.* **2002**, *36*, 218–230. [CrossRef]
37. Yang, B.; Miller-Hooks, E. Adaptive routing considering delays due to signal operations. *Transp. Res. Part B Methodol.* **2004**, *38*, 385–413. [CrossRef]
38. Gao, S.; Chabini, I. Optimal routing policy problems in stochastic time-dependent networks. *Transp. Res. Part B Methodol.* **2006**, *40*, 93–122. [CrossRef]
39. Nielsen, L.R.; Andersen, K.A.; Pretolani, D. Ranking paths in stochastic time-dependent networks. *Eur. J. Oper. Res.* **2014**, *236*, 903–914. [CrossRef]
40. Morandi, V.; Peirano, L.; Speranza, M.G. Real-time rerouting of traffic flows to prevent disruptions. In Proceedings of the TRISTAN XII Symposium, Okinawa, Japan, 22–27 June 2025.
41. Morandi, V. Bridging the user equilibrium and the system optimum in static traffic assignment: A review. *4OR* **2024**, *22*, 89–119. [CrossRef]
42. Muñoz-Villamizar, A.; Solano-Charris, E.; AzadDisfany, M.; Reyes-Rubiano, L. Study of urban-traffic congestion based on Google Maps API: The case of Boston. *IFAC-PapersOnLine* **2021**, *54*, 211–216. [CrossRef]
43. Kumar Gannina, A.R.; Jaffarullah, A.A.; Reddy, T.M.; Subba Reddy, S.M.; Vikas, A.S.; Mathi, S.; Ramalingam, V. A New Approach to Road Incident Detection Leveraging Live Traffic Data: An Empirical Investigation. *Procedia Comput. Sci.* **2024**, *235*, 2288–2296. [CrossRef]
44. Elnaggar, G.R.; Al-Hourani, S.; Abutaha, R. Real-time urban congestion monitoring in Jeddah, Saudi Arabia, using the Google Maps API: A data-driven framework for Middle Eastern cities. *Sustainability* **2025**, *17*, 8194. [CrossRef]
45. Rouky, N.; Bousouf, A.; Benmoussa, O.; Fri, M. A spatiotemporal analysis of traffic congestion patterns using clustering algorithms: A case study of Casablanca. *Decis. Anal. J.* **2024**, *10*, 100404. [CrossRef]
46. Amin-Naseri, M.; Chakraborty, P.; Sharma, A.; Gilbert, S.B.; Hong, M. Evaluating the Reliability, Coverage, and Added Value of Crowdsourced Traffic Incident Reports from Waze. *Transp. Res. Rec.* **2018**, *2672*, 34–43. [CrossRef]
47. Zhang, Z.; Liu, Y.; Han, L.D.; Freeze, P.B. Secondary Crash Identification using Crowdsourced Waze User Reports. *Transp. Res. Rec.* **2021**, *2675*, 853–862. [CrossRef]
48. Harsha, K.K. Exploring the Role of IOT Sensors in Enhancing Urban Mobility. *Int. Res. J. Adv. Eng. Hub* **2024**, *2*, 1181–1186. [CrossRef]
49. Varjola, M.; Loffler, J. PRONTO: Event recognition for public transport. In Proceedings of the 17th ITS World Congress, Busan, Republic of Korea, 25–29 October 2010; pp. 39–47.
50. Ichoua, S.; Gendreau, M.; Potvin, J.Y. Vehicle dispatching with time-dependent travel times. *Eur. J. Oper. Res.* **2003**, *144*, 379–396. [CrossRef]
51. Ghiani, G.; Guerriero, E. A Note on the Ichoua, Gendreau, and Potvin (2003) Travel Time Model. *Transp. Sci.* **2014**, *48*, 458–462. [CrossRef]
52. Malandraki, C.; Daskin, M.S. Time Dependent Vehicle Routing Problems: Formulations, Properties and Heuristic Algorithms. *Transp. Sci.* **1992**, *26*, 185–200. [CrossRef]
53. Hashemi, M.; Karimi, H.A. A weight-based map-matching algorithm for vehicle navigation in complex urban networks. *J. Intell. Transp. Syst.* **2016**, *20*, 573–590. [CrossRef]
54. Gmira, M.; Gendreau, M.; Lodi, A.; Potvin, J.Y. Travel speed prediction based on learning methods for home delivery. *EURO J. Transp. Logist.* **2020**, *9*, 100006. [CrossRef]
55. Vidal, T.; Martinelli, R.; Pham, T.A.; Hà, M.H. Arc Routing with Time-Dependent Travel Times and Paths. *Transp. Sci.* **2021**, *55*, 706–724. [CrossRef]
56. Foschini, L.; Hershberger, J.; Suri, S. On the Complexity of Time-Dependent Shortest Paths. *Algorithmica* **2014**, *68*, 1075–1097. [CrossRef]
57. Gmira, M.; Gendreau, M.; Lodi, A.; Potvin, J.Y. Tabu search for the time-dependent vehicle routing problem with time windows on a road network. *Eur. J. Oper. Res.* **2021**, *288*, 129–140. [CrossRef]
58. Cordeau, J.F.; Ghiani, G.; Guerriero, E. Analysis and Branch-and-Cut Algorithm for the Time-Dependent Travelling Salesman Problem. *Transp. Sci.* **2014**, *48*, 46–58. [CrossRef]
59. Adamo, T.; Ghiani, G.; Guerriero, E. An enhanced lower bound for the Time-Dependent Travelling Salesman Problem. *Comput. Oper. Res.* **2020**, *113*, 104795. [CrossRef]
60. Adamo, T.; Ghiani, G.; Guerriero, E. On path ranking in time-dependent graphs. *Comput. Oper. Res.* **2021**, *135*, 105446. [CrossRef]

61. Sharma, G. Web Crawling and Scraping: A Survey. In Proceedings of the 2024 International Conference on Healthcare Innovations, Software and Engineering Technologies (HISET), Karad, India, 18–19 January 2024; pp. 190–192. [CrossRef]
62. Gu, Y.; Qian, Z.S.; Chen, F. From Twitter to detector: Real-time traffic incident detection using social media data. *Transp. Res. Part C Emerg. Technol.* **2016**, *67*, 321–342. [CrossRef]
63. Salas, A.; Georgakis, P.; Nwagboso, C.; Ammari, A.; Petalas, I. Traffic event detection framework using social media. In Proceedings of the 2017 IEEE International Conference on Smart Grid and Smart Cities (ICSGSC), Singapore, 23–26 July 2017; pp. 303–307. [CrossRef]
64. Alomari, E.; Mehmood, R.; Katib, I. Road Traffic Event Detection Using Twitter Data, Machine Learning, and Apache Spark. In Proceedings of the 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Leicester, UK, 19–23 August 2019; pp. 1888–1895. [CrossRef]
65. Alomari, E.; Katib, I.; Mehmood, R. Iktishaf: A Big Data Road-Traffic Event Detection Tool Using Twitter and Spark Machine Learning. *Mob. Netw. Appl.* **2023**, *28*, 603–618. [CrossRef]
66. Suat-Rojas, N.; Gutierrez-Osorio, C.; Pedraza, C. Extraction and Analysis of Social Networks Data to Detect Traffic Accidents. *Information* **2022**, *13*, 26. [CrossRef]
67. Wan, X.; Lucic, M.C.; Ghazzai, H.; Massoud, Y. Empowering Real-Time Traffic Reporting Systems with NLP-Processed Social Media Data. *IEEE Open J. Intell. Transp. Syst.* **2020**, *1*, 159–175. [CrossRef]
68. Khatri, C. Real-time Road Traffic Information Detection Through Social Media. *arXiv* **2018**, arXiv:1801.05088. [CrossRef]
69. Wang, D.; Al-Rubaie, A.; Clarke, S.S.; Davies, J. Real-Time Traffic Event Detection from Social Media. *ACM Trans. Internet Technol.* **2017**, *18*, 9. [CrossRef]
70. Putra, P.K.; Mahendra, R.; Budi, I. Traffic and road conditions monitoring system using extracted information from Twitter. *J. Big Data* **2022**, *9*, 65. [CrossRef]
71. Salas, A.; Georgakis, P.; Petalas, Y. Incident detection using data from social media. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017; pp. 751–755. [CrossRef]
72. Chang, H.; Li, L.; Huang, J.; Zhang, Q.; Chin, K.S. Tracking traffic congestion and accidents using social media data: A case study of Shanghai. *Accid. Anal. Prev.* **2022**, *169*, 106618. [CrossRef]
73. Babbar, S.; Bedi, J. Real-time traffic, accident, and potholes detection by deep learning techniques: A modern approach for traffic management. *Neural Comput. Appl.* **2023**, *35*, 19465–19479. [CrossRef]
74. Aburas, H.; Shahrour, I.; Sadek, M. Leveraging Crowdsourcing for Mapping Mobility Restrictions in Data-Limited Regions. *Smart Cities* **2024**, *7*, 2572–2593. [CrossRef]
75. Sabah Mredula, M.; Rahman, M.S.; Hosen, A.S.M.S. Accident event detection from Facebook posts written in Bengali and Banglish languages. *Int. J. Commun. Syst.* **2025**, *38*, e5671. [CrossRef]
76. Olteanu, A.; Castillo, C.; Diaz, F.; Kıcıman, E. Social data: Biases, methodological pitfalls, and ethical boundaries. *Front. Big Data* **2019**, *2*, 13. [CrossRef]
77. Kitchin, R. The ethics of smart cities and urban science. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20160115. [CrossRef]
78. Cottrill, C.D. MaaS surveillance: Privacy considerations in mobility as a service. *Transp. Res. Part A Policy Pract.* **2020**, *131*, 50–57. [CrossRef]
79. Liao, C.; Brown, D.; Fei, D.; Long, X.; Chen, D.; Che, S. Big data-enabled social sensing in spatial analysis: Potentials and pitfalls. *Trans. GIS* **2018**, *22*, 1351–1371. [CrossRef]
80. van Hulst, J.M.; Zeni, M.; Kröller, A.; Moons, C.; Casale, P. Beyond Privacy Regulations: An Ethical Approach to Data Usage in Transportation. *arXiv* **2020**, arXiv:2004.00491. [CrossRef]
81. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P.S. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 4–24. [CrossRef]
82. Qi, X.; Yao, J.; Wang, P.; Shi, T.; Zhang, Y.; Zhao, X. Combining weather factors to predict traffic flow: A spatial-temporal fusion graph convolutional network-based deep learning approach. *IET Intell. Transp. Syst.* **2024**, *18*, 528–539. [CrossRef]
83. Huang, X.; Jiang, Y.; Wang, J.; Lan, Y.; Chen, H. A multi-modal attention neural network for traffic flow prediction by capturing long-short term sequence correlation. *Sci. Rep.* **2023**, *13*, 21859. [CrossRef] [PubMed]
84. Lian, Q.; Sun, W.; Dong, W. Hierarchical Spatial-Temporal Neural Network with Attention Mechanism for Traffic Flow Forecasting. *Appl. Sci.* **2023**, *13*, 9729. [CrossRef]
85. Zhang, Z.; Li, M.; Lin, X.; Wang, Y.; He, F. Multistep Speed Prediction on Traffic Networks: A Graph Convolutional Sequence-to-Sequence Learning Approach with Attention Mechanism. *arXiv* **2018**, arXiv:1810.10237. [CrossRef]
86. Zhou, Y. An urban traffic flow prediction method based on multi-source data fusion. In Proceedings of the 2025 International Conference on Software Engineering and Computer Applications, Kunming, China, 13–15 June 2025; pp. 308–313. [CrossRef]

87. Wang, J.; Hu, Y. Enhancing spatial and textual analysis with EUPEG: An extensible and unified platform for evaluating geoparsers. *Trans. GIS* **2019**, *23*, 1393–1419. [CrossRef]
88. Wang, J.; Hu, Y.; Joseph, K. NeuroTPR: A neuro-net toponym recognition model for extracting locations from social media messages. *Trans. GIS* **2020**, *24*, 719–735. [CrossRef]
89. Idakwo, P.O.; Adekanmbi, O.; Soronnadi, A.; David, A. Geo-parsing and analysis of road traffic crash incidents for data-driven emergency response planning. *Heliyon* **2025**, *11*, e41067. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.