

Article

A Multi-Source Machine Learning Framework for Segment-Level Travel Time Prediction in Urban Arterial Corridors: Toward Sustainable Traffic Management

Muhammed Enes Karaoglan ¹ and Yetis Sazi Murat ^{2,*}

¹ The Graduate School of Natural and Applied Sciences, Department of Civil Engineering, Pamukkale University, Denizli 20160, Turkey

² Department of Civil Engineering, Faculty of Engineering, Pamukkale University, Denizli 20160, Turkey

* Correspondence: ysmurat@pau.edu.tr

Abstract

Accurate short-term travel time prediction is foundational for sustainable urban mobility and intelligent transportation systems on urban arterial corridors, where travel conditions are shaped by interacting traffic, weather, and public transport factors. This study proposes a multi-source machine learning framework for segment-direction-level prediction in the Denizli city center. Floating car data (FCD), Traffic Control Center (TCC) inductive loop detector measurements, historical weather, and public transport indicators were integrated into a 15 min time-segment structure. The final dataset includes 60 segment-direction targets. Performance was evaluated using Linear Regression, Random Forest, LightGBM, and LSTM under a chronological train-validation-test design. Tree-based ensemble models produced the most stable overall performance, with LightGBM and Random Forest yielding similarly low pooled test errors. Segment-level analyses revealed clear spatial and temporal heterogeneity, showing no single model is universally superior across all links. By providing reliable traffic-state information, the framework enables efficient traffic management and may indirectly reduce delay, fuel use, and emissions; these environmental effects were not quantified. SHAP-based interpretation showed that temporal and traffic-state variables dominate predictions, while weather and public transport provide complementary value.

Keywords: sustainable urban mobility; floating car data; machine learning; public transport effect; traffic state; travel time prediction; urban arterials



Academic Editors: Vincenzo Gallelli and Rosolino Vaiana

Received: 3 July 2026

Revised: 23 July 2026

Accepted: 30 July 2026

Published: 7 August 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Short-term travel time prediction is a practical requirement for urban traffic management because travel time uncertainty directly affects route guidance, traveler information, signal coordination, and network monitoring [1]. In medium-sized cities, this requirement is complicated by fragmented data infrastructures: travel times, detector measurements, weather records, and public transport information are often generated by different systems and must be harmonized before they can support operational prediction.

On urban arterials, travel time is shaped by nonlinear interactions among traffic volume, speed, capacity, weather, incidents, signal control, stops, and demand fluctuations [2,3]. These interactions are often local rather than corridor-wide, so prediction models need to represent both temporal variation and segment-level spatial heterogeneity.

These conditions motivate a modeling unit that is finer than the corridor average and a data pipeline that can combine heterogeneous observations at a common time-space resolution.

Machine learning methods are suitable for this setting because they can learn non-linear patterns from high-dimensional multi-source data without requiring predefined functional forms [4,5]. However, their usefulness for traffic management depends not only on aggregate accuracy but also on whether model reliability can be interpreted at the segment level.

Much of the existing travel time prediction literature still emphasizes freeways, large metropolitan areas, or corridor/link-level outputs [5,6]. For arterial corridors, adjacent segments may behave differently because of intersection density, signal operation, lane configuration, stop activity, and public transport interactions; corridor averages can therefore conceal local delay mechanisms.

Integrated segment-level studies that jointly consider FCD, TCC loop detector measurements, meteorological variables, and public transport exposure are particularly limited for medium-sized cities. This study addresses that gap by focusing on data fragmentation and segment-direction-level prediction in Denizli, a non-metropolitan but data-rich urban network.

Accordingly, three connected methodological gaps remain. First, single-source and corridor-averaged models do not adequately represent local arterial heterogeneity. Second, studies that fuse fixed and mobile traffic observations rarely incorporate meteorology and public transport exposure within the same operational table. Third, strong pooled accuracy can conceal segment-directions for which predictions are unreliable. The proposed framework addresses these gaps through common-resolution data fusion, independent segment-direction models, chronological validation, and local reliability diagnostics.

The main motivation of this study is to address these research gaps by developing a segment-level travel time prediction framework based on multi-source data integration. The following research questions are examined:

- Can short-term segment-direction travel time be reliably predicted using multi-source urban traffic data?
- Do meteorological variables and indicators representing the public transport effect make a meaningful contribution to travel time prediction performance?
- How does prediction performance vary across corridors, directions, and traffic segments?

This study presents a comprehensive application on segments located on six main arterial corridors in the Denizli city center. The target variable is segment-direction travel time, while the explanatory variables consist of temporal indicators, TCC inductive loop detector measurements, meteorological data, and variables representing the public transport effect. The analyses utilized a 15 min temporal resolution, which captures short-term traffic changes while preserving model stability.

The novelty of this study is methodological and application-oriented, rather than focusing on the development of a new learning algorithm. It lies in jointly harmonizing four institutionally fragmented data sources at the segment-direction level, representing public transport exposure through dynamic PCU-based indicators, and evaluating both predictive accuracy and local reliability under a leakage-resistant chronological protocol. This combination distinguishes the framework from studies that focus on a single source, freeway-scale states, route averages, or a single aggregate model score.

The main contributions of this study are summarized as follows:

The study develops a multi-source data pipeline to harmonize FCD, TCC loop detector, meteorological, and public transport exposure data in a common 15 min segment-direction structure.

Short-term travel time is modeled independently for the retained segment-directions, enabling local performance differences that would be hidden by corridor-level averages to be evaluated.

Linear Regression, Random Forest, LightGBM, and LSTM models are compared under a chronological train-validation-test design, using nonparametric statistical testing to avoid over-interpreting small aggregate differences.

The results are interpreted through segment-direction error profiles and SHAP-based feature-group analysis, linking prediction accuracy to operational traffic-management use rather than reporting only pooled model scores.

Together, these contributions connect multi-source data integration, local error diagnosis, and decision support for medium-sized urban arterial corridors.

The remainder of the paper is organized as follows. Section 2 reviews the literature on short-term travel time prediction and machine learning applications. Section 3 describes the study area and dataset. Section 4 presents the data preparation and integration processes. Section 5 defines the feature engineering procedure and modeling variables. Section 6 explains the methodology and experimental design. Section 7 presents the model performance results, Section 8 discusses the findings, and the final section presents the conclusions.

2. Literature Review

Studies on travel time prediction have undergone a substantial transition from early approaches based on traffic flow data collected from fixed sensors to multi-source and data-driven modeling frameworks [7]. Early studies were primarily conducted on freeways and access-controlled road networks, where variables such as traffic volume, speed, and occupancy were used for short-term traffic-state and travel time estimation [8]. Later studies moved toward data-driven and machine learning-based methods that account for more complex processes on urban arterials, including signal control, intersection delays, stop effects, weather conditions, and demand fluctuations [9].

The literature addresses travel time prediction at different spatial scales [10]. Corridor-, link-, route-, and trip-level studies provide different contributions to traffic management and traveler information systems [11]. However, traffic conditions on urban arterials are not homogeneous along a corridor; travel time behavior can vary even across nearby segments because of intersection layout, signal operation, stop locations, and local demand differences [12]. More detailed approaches that consider links, segments, and trip sub-components have therefore gained attention in recent years [13]. Nevertheless, studies on traffic-state estimation for segments with missing data in urban road networks remain an active area of research [14].

The data sources used in the travel time prediction literature have become increasingly diverse with technological advances [15]. Traditional approaches largely rely on speed, volume, and occupancy data obtained from inductive loop detectors and similar fixed infrastructure components. In contrast, the spread of GPS- and navigation-based systems has systematically increased the use of FCD in recent studies. FCD has become a strategic data source because it enables direct observation of travel times and provides extensive data coverage across broad areas of urban networks. However, the low-frequency nature of FCD, data sparsity, and positional uncertainty introduce methodological challenges in map matching and path inference, requiring a careful data processing workflow. Although some studies argue that GPS data alone may be sufficient for reliable predictions, combining these data with meteorological information and signal-control parameters in hybrid models can substantially improve prediction sensitivity under unpredictable traffic conditions [16]. Many methods in the literature also focus on detailed processing of spatial and temporal data to maintain prediction quality under different traffic conditions [17].

Using different data sources together aims to represent traffic patterns that cannot be explained by a single source. When fixed sensors, FCD, meteorological indicators, incident information, and network features are used jointly, both the temporal and spatial

dimensions of traffic state can be represented more strongly. However, multi-source data integration requires a separate data preparation and quality control process because of different temporal resolutions, missing observations, spatial matching errors, and inconsistencies across sources.

Although traffic flow variables are the primary inputs for travel time prediction, external factors can also contribute to model performance. Several studies have shown the effects of weather conditions on traffic speed, capacity, reliability, and travel time [18]. Variables such as rain and snow have been reported to affect macroscopic travel times in urban networks [19]. Including meteorological variables in travel time prediction models is therefore important for representing deviations from normal traffic patterns, especially in short-term prediction.

Temporal variables are also widely used in short-term prediction problems [20]. Time of day, day of week, weekday/weekend status, and intra-day periods represent recurrent demand patterns and help the model learn the basic traffic rhythm. Although these variables do not explain all traffic variability by themselves, they play a complementary role in modeling regular peak-hour behavior and weekly demand cycles.

Studies that directly include the effect of public transport activity on general traffic travel time in prediction models are more limited [21]. Much of the existing literature focuses on bus arrival time, bus travel time, or AVL-based public transport performance prediction [22]. These studies show that public transport data can be used in operational prediction, but variables representing the effect of bus and minibus movements on general mixed-traffic segment travel time remain less developed. Researchers use passenger car unit (PCU) approaches to convert the effect of different vehicle types on mixed traffic flow into a common unit. In addition, processing AVL and GPS data with geofencing techniques to obtain stop-based link times provides a critical mechanism for analyzing the traffic flow dynamics of public transport vehicles on urban arterials [23].

Methodologically, travel time prediction studies have shifted from linear models and ARIMA-based time-series approaches toward machine learning and deep learning methods [24]. Support Vector Regression stood out in early machine learning applications because it can model nonlinear relationships. Later, tree-based ensemble methods such as Random Forest, Gradient Boosting, XGBoost, and LightGBM 4.6.0 became strong options for tabular and multi-source traffic data [25].

Deep learning-based approaches have also become common in the traffic prediction literature. LSTM-based models have been used in short-term speed and traffic flow prediction because they can learn long-term dependencies in traffic time series [26]. More recent studies jointly handle spatial and temporal dependencies using graph neural networks, attention mechanisms, and end-to-end travel time prediction architectures. However, it is also reported that tree-based ensemble methods can produce results similar to, and in some cases stronger than, those of deep learning models on structured tabular datasets [27].

The model validation strategy has particular importance in traffic time-series studies. Because traffic data contain temporal dependence, randomly splitting training, validation, and test sets can cause information leakage and unrealistic performance estimates. Accordingly, preserving the temporal structure of the data is essential in time-series prediction problems, and chronological data partitioning and rolling-origin validation strategies are recommended [28]. Reporting multiple error metrics such as MAE, RMSE, and MAPE also enables a comprehensive model analysis by separating the absolute and relative effects of error magnitude.

Finally, machine learning models are increasingly evaluated not only in terms of prediction accuracy but also in terms of interpretability and statistical comparison. SHAP-based explanations are widely used to assess the relative contribution of variables to model

predictions, especially in tree-based models [29]. When multiple models are compared on the same data partitions, paired nonparametric tests and multiple-comparison corrections are recommended.

Recent studies provide complementary but incomplete responses to these challenges. Hyperparameter-optimized classifiers have examined short-term freeway traffic states at different time intervals [30], and quantile regression forests have analyzed short-term freeway speed prediction across 5, 10, and 15 min horizons [31], while a recent arterial study fused loop detector and probe-vehicle data through predefined traffic-flow sections [32]. Other work has emphasized temporally correlated feature selection [10] or optimized trajectory granularity to balance spatial detail and computational efficiency [33]. These advances demonstrate the value of optimization, fusion, temporal structure, and spatial representation, but they do not jointly address FCD, loop detector, meteorological, and public transport information at the segment-direction level with chronological out-of-sample testing and local reliability analysis. This unresolved combination defines the methodological gap addressed here. Furthermore, the success of data-driven calibration and advanced optimization algorithms is not limited to transportation networks; these machine learning frameworks are increasingly demonstrating their critical effectiveness in managing nonlinearities across other complex engineering domains, such as the kinematic calibration of industrial manipulators [34,35].

Microsimulation-based optimization approaches in signalized intersection and corridor management have been shown to improve delay, capacity utilization, and emission indicators [36]. In contrast, the present study focuses on segment-direction-level travel time estimation with a machine learning framework. By defining physical segments, treating out-bound and inbound flows as separate prediction targets, and integrating multi-source data, the proposed approach accounts for heterogeneous traffic dynamics and public transport interactions that can be overlooked in aggregated corridor-level models.

3. Study Area and Dataset Description

3.1. Study Area

The application area of this study comprises the main urban transportation corridors in the Denizli city center, Türkiye. Although Denizli is not a large metropolitan area, it represents a medium-sized city with intensive arterial use, signalized intersections, mixed traffic, and public transport movements in the city center. Denizli was selected as the study area because it has TCC infrastructure, inductive loop detector data are available, and FCD can provide complementary network coverage. These characteristics make Denizli a suitable setting for evaluating short-term travel time prediction in a data-rich but non-metropolitan urban context.

In this study, six main corridors representing traffic flow in the Denizli city center were considered (Figure 1): Adliye–Tiyatro, Tiyatro–Ulus, Ulus–Kelleci, Hal–Hallaclar, Hallaclar–Adliye, and YSE–Cinar. These corridors were selected to cover arterial traffic in different directions, intersection passages, and route sections where public transport interaction can be observed. Each corridor was divided into segments based on physical intersections and nodes, and outbound and inbound directions were evaluated separately.

Figure 1 shows the study area and the six corridors evaluated in the analysis. Each corridor is represented by a different color on the map, and the intersections and nodes forming the segment boundaries are numbered. The names of the numbered points are provided in the legend so that corridor routes and the segment structure can be read together.

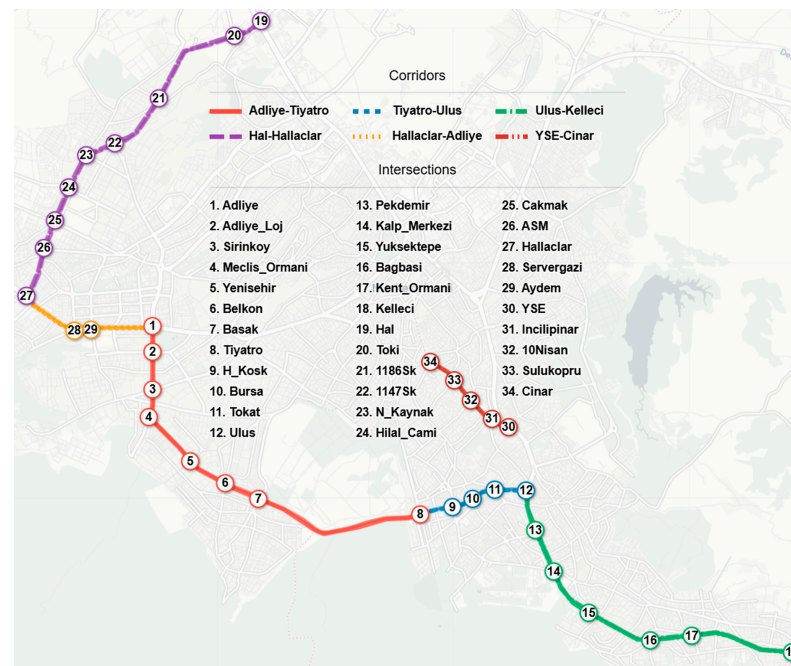


Figure 1. Map representation of the Denizli study area and the six corridors.

3.2. Corridor and Segment Definition

The dataset was structured to support segment-direction-level prediction rather than corridor-level prediction. This choice avoids assuming homogeneous traffic flow along an urban corridor and enables local delay patterns to be modeled separately.

In this study, a physical segment is defined as the route section between two consecutive intersections or nodes numbered on the map. Figure 2 shows the schematic representation of the segment-direction structure. Because outbound and inbound directions are treated as separate target variables, 64 segment-direction targets were obtained from 32 physical segments.

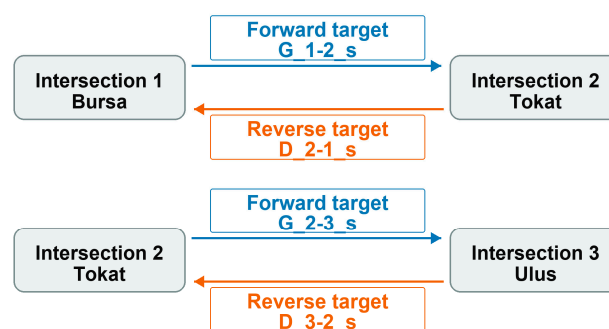


Figure 2. Schematic representation of the segment-direction structure.

The target variables were defined in the Excel-based master data template in the format 'Target:.._s'. The prefix G_ denotes the outbound direction, the prefix R_ denotes the inbound direction, the segment name indicates the corresponding start and end nodes, and the suffix _s indicates that travel time is expressed in seconds. For example, the variable Target:G_Tokat_Ulus_s represents the travel time target, in seconds, for the segment from Tokat Intersection toward Ulus Intersection.

Figure 2 illustrates that the same physical road section may have different traffic conditions in outbound and inbound directions and is therefore modeled as two separate prediction targets.

Table 1 presents the number of physical segments and the resulting segment-direction targets for each corridor. As shown, the number of segment-direction targets is twice the number of physical segments in each corridor. This structure makes it possible to model direction-specific traffic conditions separately.

Table 1. Corridor, segment, and target variable structure.

Corridor	Physical Segment Count	Segment-Direction Target Count	Physical Segments in Outbound Direction
Adliye–Tiyatro	7	14	Adliye–Adliye_Loj; Adliye_Loj–Sirinkoy; Sirinkoy–Meclis_Ormani; Meclis_Ormani–Yenisehir; Yenisehir–Belkon; Belkon–Basak; Basak–Tiyatro
Tiyatro–Ulus	4	8	Tiyatro–H_Kosk; H_Kosk–Bursa; Bursa–Tokat; Tokat–Ulus
Ulus–Kelleci	6	12	Ulus–Pekdemir; Pekdemir–Kalp_Merkezi; Kalp_Merkezi–Yuksektepe; Yuksektepe–Bagbasi; Bagbasi–Kent_Ormani; Kent_Ormani–Kelleci
Hal–Hallaclar	8	16	Hal–Toki; Toki–1186Sk; 1186Sk–1147Sk; 1147Sk–N_Kaynak; N_Kaynak–Hilal_Cami; Hilal_Cami–Cakmak; Cakmak–ASM; ASM–Hallaclar
Hallaclar–Adliye	3	6	Hallaclar–Servergazi; Servergazi–Aydem; Aydem–Adliye
YSE–Cinar	4	8	YSE–Incilipinar; Incilipinar–10Nisan; 10Nisan–Sulukopru; Sulukopru–Cinar
Total	32	64	-

3.3. Data Sources and Collection Strategy

A multi-source data structure was used in this study. The main dataset consists of FCD-based travel time observations, inductive loop detector data, historical weather data, and public transport exposure/passenger car unit (PTE/PCU) variables. A commercial traffic API provided FCD-based travel time observations [37], and a historical weather API provided meteorological records [38]. All data sources were merged on a common time axis and structured at the segment-direction level.

FCD was used to obtain segment-based travel time information. During data collection, outbound and inbound routes of the corridors were defined using start, end, and intermediate points. The resulting travel time values were evaluated together with segment lengths to derive segment-based speed information. Researchers applied a more frequent query strategy during daytime hours and a less frequent strategy at night. All observations were then converted to a common 15 min temporal resolution. FCD-derived speed variables were not used as independent TCC inductive loop detector variables in the final modeling stage; they served mainly as auxiliary inputs in the PTE/PCU calculations.

TCC data include speed, volume, and occupancy measurements collected through inductive loop detectors connected to the Denizli Metropolitan Municipality Traffic Control Center network. Different data sources were integrated into a common structure through matching dictionaries created between detector labels and the segment variables in the master data template. Some detectors experienced long-term data gaps or failures during the study period; these detector-availability problems were handled in the quality-control procedures described in Sections 4.2 and 4.4.

Meteorological data were obtained from regional locations representing the study area. The variables used were weather code, precipitation amount, temperature, and visibility.

Hourly meteorological records were converted to the 15 min time structure to align them with the other data sources.

PTE/PCU variables derived from bus and minibuses movements were used to represent the public transport effect. GPS data were evaluated for buses, while line, stop, and frequency information were evaluated for minibuses. The resulting values were calculated by segment for 15 min time intervals and then converted into a common traffic-load indicator using the PCU approach. The PCU conversion method and calibration process are described in detail in Section 4.

3.4. Target Variables and Explanatory Variable Groups

The prediction target in this study is travel time in seconds for each segment-direction. Target variables were obtained from FCD and defined in the master dataset in the ‘Target:.._s’ format. The final modeling scope after detector-availability quality control is reported in Section 4.4 and summarized in Table 2.

Table 2. Scope of this study.

Scope Element	Description
Study area	Denizli city center, Türkiye
Study period	5 January 2026 to 14 May 2026
Temporal resolution	15 min
Number of corridors	6
Number of physical segments	32
Raw segment-direction target count	64
Number of segment-directions included in machine learning	60
Number of segment-directions excluded from modeling	4: segment-directions for which TCC/inductive loop detector data could not be obtained throughout the study period
Target variable	Segment-direction travel time in seconds
Data sources	FCD, TCC/inductive loop detector data, historical weather data, and public transport PTE/PCU variables
Final modeling scenario	Common 31-variable base schema without target history or engineered lag/rolling columns; model-specific input representation is described in Sections 5.4 and 6.4

The explanatory variables were grouped under four main categories:

- Temporal variables, including date, hour, day of week, intra-day period, and weekday indicators;
- Meteorological variables;
- TCC-based speed, volume, and occupancy variables;
- PTE/PCU variables representing the public transport effect.

This structure facilitated the evaluation of traffic, environmental, and operational factors that may affect travel time within the same modeling framework.

3.5. Temporal Resolution and Observation Structure

The final dataset was structured at a 15 min temporal resolution. This interval was selected as the common operational unit because it preserves within-day peak transitions while aggregating sufficient FCD, detector, and public transport observations to limit sparsity and source-specific noise. Five- to fifteen-minute horizons are widely used for short-term traffic-state applications [30,31]; evidence from multi-horizon experiments also

shows that longer short-term intervals can stabilize noisy traffic measurements, although excessive aggregation may smooth abrupt events [31]. For the present fragmented sources, 15 min therefore provided a practical balance between temporal responsiveness and data completeness. Each day was divided into 96 intervals.

The study period covers 5 January 2026 to 14 May 2026. This period was selected because it was the longest common observation interval during which all data sources could be obtained simultaneously and with sufficient continuity.

Following data preparation and calibration, monthly calibrated data files were created, and segment-direction machine learning-ready datasets were generated from these files. For each retained segment-direction, the final dataset contains 12,478 observations, resulting in 748,680 segment-level observations in the final modeling structure.

Table 2 summarizes the main characteristics of the study area, data sources, segment structure, and final modeling scope, with the retained modeling set defined after the quality-control procedure explained in Section 4.4.

4. Data Preparation and Calibration

This section explains the data integration, missing-data handling, quality-control, and calibration procedures used to make the multi-source dataset introduced in Section 3 suitable for machine learning modeling. The main purpose of data preparation is to merge raw observations obtained from different sources into a common time–space structure and to create a consistent and reliable segment-direction-level modeling dataset.

4.1. Multi-Source Data Integration and Temporal Alignment

The dataset used in this study consists of FCD, TCC inductive loop detector data, meteorological data, and public transport operation data. Because these sources have different temporal resolutions, recording structures, and spatial reference systems, they had to be transformed into a common data structure before modeling.

Data integration was performed in two main stages. First, all data sources were transferred to a common time axis and reorganized using 15 min time intervals. Second, each data source was matched with the segment-direction structure, and all variables were combined in a common observation table. Thus, each row was structured to represent traffic, meteorological, and public transport conditions for a specific segment-direction and a specific 15 min time interval.

FCD provided segment-based travel time information, while TCC data included speed, volume, and occupancy measurements. Meteorological data were adapted from hourly records to a 15 min temporal resolution, and public transport data were transferred to the common time structure for use in the subsequent feature engineering stage. Data fusion was therefore treated not simply as merging different files, but as harmonizing different data-generation logics within the same modeling structure. Figure 3 summarizes the main processing steps from the transformation of raw data sources into a common time-segment structure to the creation of the final modeling dataset.

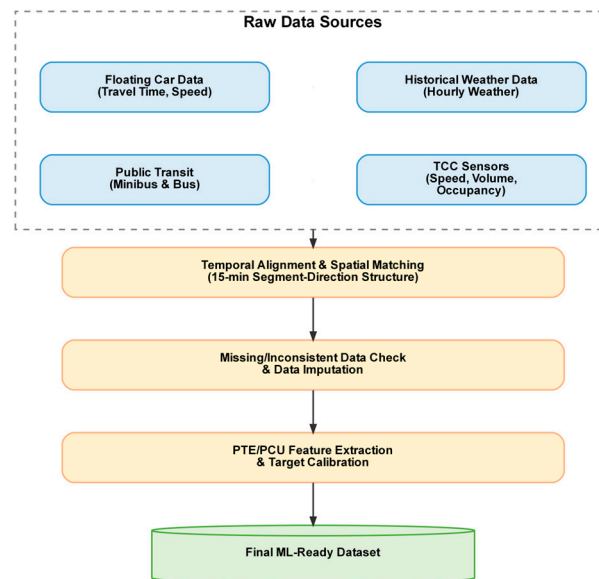


Figure 3. General workflow of the data preparation and calibration process.

4.2. Missing Data Management and Quality Control

After data integration, all data sources were evaluated in terms of missing observations, temporal consistency, and data integrity. Missing records may result from temporary service outages at data providers, communication interruptions, inductive loop detector failures, database issues, or the absence of generated data in the relevant time interval.

Missing data management in this study was not performed using a single uniform method for all variables. Instead, a variable-specific approach was adopted by considering the data-generation logic of each source.

Missingness was handled separately by source and variable type. Each TCC speed, volume, and occupancy series was first sorted chronologically. Short, isolated gaps of up to two consecutive 15 min intervals were filled by time-based linear interpolation between the nearest valid observations. Extended or otherwise unresolved gaps were filled from the variable's historical weekly profile, defined as the arithmetic mean of all available valid observations for the same detector, day of week, and 15 min clock time; no fixed number of reference weeks was imposed. If that cell-specific profile was unavailable, the median of the valid observations in the same detector series was used as a final numerical fallback. Values were not transferred between different detectors. Segment-directions affected by persistent detector outage or insufficient reliable coverage were excluded during final quality control rather than retained through extensive imputation, as described in Section 4.4.

Missing FCD-derived target travel times (S_n) and their paired speed fields were processed separately from TCC explanatory variables. For each segment-direction and hour, the segment-hour reference mean (avg_reference_sn) was used as the center, with a segment-level reference when the hourly value was unavailable. Relative-deviation distributions were estimated from all valid observations after the missing period, without a fixed weekly or day-type window. Hour-specific distribution parameters were used when sufficient observations were available; otherwise, segment-global parameters were used. Extreme deviations and standard deviations were bounded before seeded normal sampling, and sampled deviations were clipped before calculating $S_n^* = \max[1, \text{avg_reference_sn}(1 + d^*)]$. The paired speed was then recomputed as $3.6L/S_n^*$, where L is segment length in meters, preserving distance–time consistency. This target-repair procedure was not used to mask persistent detector failure.

Missing data in meteorological records were limited; the main operation was converting hourly records into 15 min time intervals. Weather code, precipitation, temperature, and visibility variables were assigned to the four time intervals within the corresponding hour.

For public transport data, the operating logic was considered. Time intervals at night when minibuss service was not provided were treated not as missing data but as true service absence, and the corresponding observations were coded as zero.

The following criteria were systematically examined during the quality-control process:

- Missing or empty observations;
- Duplicate records;
- Timestamp inconsistencies;
- Consecutive missing time intervals;
- Physically impossible speed and travel time values;
- Inconsistencies between inductive loop detector data and target travel time;
- Incompatibility of public transport data with the operating logic.

4.3. Target Travel Time Calibration

The dependent variable, segment-direction travel time, directly affects model reliability and was therefore subjected to a targeted calibration process. To identify systematic data-generation errors rather than mask normal traffic variability, observations were compared against hourly and daily reference profiles. The quality-assurance process specifically addressed and corrected:

- Route definition errors;
- Coordinate matching problems;
- Incorrect segment assignments;
- Inconsistencies caused by data retrieval processes;
- Abnormal travel time jumps.

Importantly, high or low travel times caused by physical traffic events—such as crashes, road works, temporary capacity reductions, or exceptional congestion—were explicitly retained in the dataset. Figure 4 illustrates the correction of a coordinate definition error, demonstrating that this procedure is a controlled quality-assurance step rather than a general data-smoothing operation.

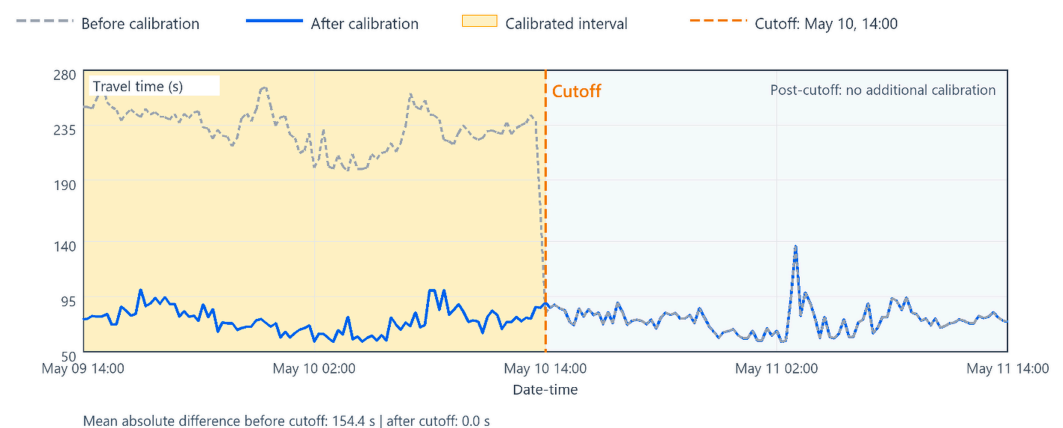


Figure 4. Example of travel time calibration after correcting a coordinate definition.

4.4. Creation of the Final Dataset

After completing data integration, missing-data management, and calibration, monthly calibrated data files were created, and segment-direction modeling datasets were generated from them.

In the final quality-control process, the following criteria were evaluated jointly before constructing the modeling files:

- Alignment of timestamps with the common 15 min structure;
- Segment-direction matches;
- Continuity of inductive loop detector data;
- Completeness of meteorological variables;
- Compatibility of public transport data with the operating logic;
- Consistency of target travel time values.

This checklist was used to determine whether each segment-direction could be retained in the machine learning dataset.

At the end of this process, four segment-directions were excluded because reliable TCC explanatory variables could not be produced throughout the study period. Consequently, 60 of the 64 raw segment-direction targets were retained for the final analyses.

Each observation in the final dataset integrates travel time, traffic state, meteorological conditions, and public transport exposure for a specific segment-direction and a specific 15 min time interval. This uniform structure enables each segment-direction to be treated as an independent prediction problem. The specific feature engineering and variable-generation methods applied to this dataset are detailed in the next section.

5. Feature Engineering and Model Input Structure

5.1. Target Variable and Modeling Unit

Each segment-direction functioned as a unique prediction task during the machine learning stage. The segment-direction travel time variables stored in the master data template in the 'Target:.._s' format were standardized in the modeling files and used as the s target variable. The s variable represents segment-direction travel time in seconds for the corresponding 15 min time interval.

The analyses used the 60 segment-directions retained after the final quality-control process. Each retained segment-direction dataset contains 12,478 observations. Thus, each model was trained and evaluated separately to perform short-term travel time prediction for a specific segment-direction.

5.2. Generation of Temporal Features

Urban traffic demand exhibits clear cyclical patterns during the day [39]. Using the hour information (0–23) as a linear value in machine learning models may cause the temporal proximity between 23:00 and 00:00 to be lost and interpreted by algorithms as a large mathematical distance. Conversely, categorical encoding of hours both expands the feature space and fails to fully reflect the structural similarity of neighboring time periods. To avoid this discontinuity and preserve the relative similarity inherent in the periodic nature of time, the hour variable was transformed into two new continuous variables using trigonometric sine and cosine functions through cyclic encoding, as shown in Equations (1) and (2) [40]:

$$t_{sin} = \sin\left(\frac{2\pi \cdot t}{24}\right) \quad (1)$$

$$t_{cos} = \cos\left(\frac{2\pi \cdot t}{24}\right) \quad (2)$$

Here, t denotes the relevant hour of the day (0–23). This transformation mathematically preserves the natural sequence and spatial relationship between the end of one day and the beginning of the next. In addition to the hour variable, the temporal feature group includes day of week, weekday/weekend status, public holiday indicators, and intra-day period classes as explanatory variables.

5.3. Public Transport Exposure Features

Representing the public transport effect directly by vehicle counts does not adequately capture the actual impact of buses and minibuses on traffic flow [41]. Therefore, public transport movements were converted into traffic-impact indicators based on the PCU approach.

This approach is consistent with the dynamic PCU literature, which aims to represent the relative effect of different vehicle types on traffic flow under mixed traffic conditions through a common unit.

The PCU conversion used in this study is an adaptation of the PCU approach to the segment-time structure proposed by Chandra and Kumar, which considers the physical projected area of vehicles and speed differences under mixed traffic conditions. The indicator is defined by the following equation [42]:

$$PCU_{i,s,t} = \frac{V_c}{V_{i,s,t}} \cdot \frac{A_i}{A_c} \quad (3)$$

Here, i represents the vehicle type, s the segment, t the time interval, V_c the reference passenger car speed on the segment, $V_{i,s,t}$ the average speed of the relevant vehicle type on segment s at time t , A_c the projected area of a passenger car, and A_i the projected area of the relevant vehicle type (bus or minibus).

GPS records were used to generate the bus variables. Bus location signals were filtered using influence areas defined around the nodes, and segment passages were identified through consecutive node crossings. The physical area A_i occupied by a bus was determined using vehicle dimension information from fleet lists. The calculated segment speed $V_{i,s,t}$ was then compared with passenger car values to produce a bus-based PCU value for each 15 min time interval.

For minibuses, direct vehicle GPS passage records were unavailable; therefore, line-stop associations and period-specific scheduled headways were combined with the contemporaneous segment FCD speed. For a segment associated with one or more minibus lines, the expected number of passages in a 15 min interval ($N_{\min}$) was calculated as follows

$$N_{\min} = 15 \sum_l \frac{1}{f_{l,p}} \quad (4)$$

where $f_{l,p}$ is the scheduled headway in minutes for line l during period p (morning 07:00–09:59, midday 10:00–15:59, evening 16:00–19:59, and the remaining service hours). Exposure was set to zero from 00:00 to 05:59. The segment passenger car speed v_{car} was read from the matched FCD speed field, and minibus speed was proxied as $v_{\min} = \max(0.80v_{car}, 5 \text{ km/h})$. Minibus PCU exposure ($PCU_{\min}$) was then calculated as follows:

$$PCU_{\min} = N_{\min} \cdot \frac{A_{\min}}{A_{car}} \cdot \frac{v_{car}}{v_{\min}} \quad (5)$$

The calculation used plan-area constants $A_{\min} = 13.5 \text{ m}^2$ and $A_{car} = 8.5 \text{ m}^2$. Values were set to zero when no associated line, positive FCD speed, or scheduled passage was available, and total public transport exposure was obtained by adding the separately GPS-derived bus and minibus PCU values. Thus, bus exposure reflects observed GPS passages, whereas minibus exposure is a schedule- and FCD-based proxy.

Exposure values obtained from bus and minibus movements were calculated at the segment level and transferred to the model as the PTE_{Bus} and PTE_{Minibus} variables. Thus, the public transport effect was evaluated not as a simple vehicle count but as an indicator representing dynamic traffic load on the segment.

5.4. Final Explanatory Variable Set

Model inputs were created from the following four main feature groups:

- Traffic condition variables ($\text{Speed}_{\text{TCC}}$, Volume , Occ);
- Meteorological variables ($\text{precipitation}_{\text{mm}}$, Heat_C , $\text{Visibility}_{\text{km}}$, and $\text{Weather}_{\text{Code}}$ indicators);
- Public transport effect variables (PTE_{Bus} , $\text{PTE}_{\text{Minibus}}$);
- Temporal variables ($\text{Hours}_{\text{Sin}}$, Hour_{Cos} , day-type indicators, and period indicators).

Neither lagged target travel times nor pre-engineered lagged-traffic and moving-average columns were included in the common 31-variable feature table. Linear Regression, Random Forest, and LightGBM used the explanatory vector observed at target interval t . LSTM used an architecture-specific sequence representation: for each target interval, rolling windows of the same base explanatory vectors over 8, 16, 24, or 32 consecutive 15 min intervals ending at t were supplied to the network. No past S_n value, recursively predicted travel time, or target-derived feature entered an LSTM window. Thus, the no-history-feature designation refers to the exclusion of target history and engineered lag or rolling columns, not to the removal of temporal context from LSTM. The framework remains a nowcasting design because the prediction is for interval t using explanatory information available up to and including t .

For the non-sequence tree models, an auxiliary feature-design sensitivity check compared the 31-variable base table with an otherwise identical no-target-lag table that additionally retained engineered traffic lags ($t - 1$ and $t - 2$) and one-hour moving averages. Removing these historical derivatives changed pooled Random Forest MAE from 6.926 to 6.824 s (-0.102 s) and LightGBM MAE from 7.055 to 7.089 s ($+0.034$ s). The small, mixed-direction changes (less than 0.11 s) indicate that engineered traffic-history columns did not provide a consistent aggregate advantage for these reference models. This check is distinct from the LSTM sequence representation, which arranges the unchanged base variables into rolling windows without adding target-history or engineered lag columns.

The final feature matrix contains 31 explanatory variables, excluding the target variable and monitoring fields.

5.5. Preliminary Examination of Feature Groups

Before modeling, the relationships between explanatory variable groups and target travel time were examined using the Spearman correlation coefficient. The distributions of absolute correlation values calculated for each segment-direction are shown in Figure 5.

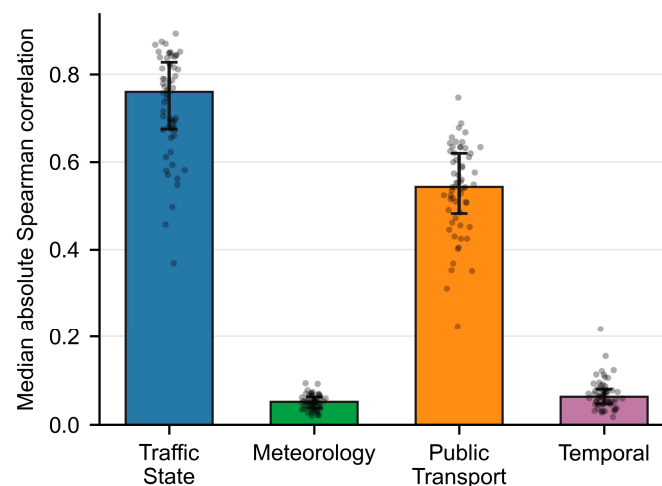


Figure 5. Relationships between target travel time and the main explanatory variable groups.

In the figure, the bars show the median absolute Spearman correlations between the explanatory variables in the relevant variable group and the target variable for each segment-direction. Error bars represent the interquartile range, and gray points represent values at the segment-direction level.

Figure 5 shows that the strongest preliminary relationship with target travel time appears in the traffic-state variables. Public transport effect variables also carry a clear explanatory signal, whereas the individual correlation levels of meteorological and temporal variables remain lower. However, these results reflect only individual relationships among variables and do not capture nonlinear relationships or interactions that machine learning models may learn.

The wide segment-direction spread in Figure 5 indicates that marginal monotonic associations are not spatially uniform. This result is complementary to, rather than inconsistent with, the SHAP analysis in Section 7.6: Spearman correlation summarizes one-variable monotonic associations, whereas SHAP measures how a fitted nonlinear model uses variables conditionally and in interaction with the remaining feature set. The two analyses are therefore interpreted jointly when distinguishing recurrent temporal structure from real-time traffic-state adjustments.

6. Methodology and Experimental Design

6.1. Problem Formulation

This study is formulated as the problem of predicting segment-direction travel time at 15 min intervals. Because the target variable is continuous, the problem is treated as a regression task. For each segment-direction, the target variable represents the observed travel time in seconds for the relevant time interval.

The modeling problem was defined separately for each segment-direction. This approach assumes that each segment-direction may have different traffic characteristics, geometric properties, signalization conditions, and public transport interactions. Let $x_{s,t}$ denote the 31-variable base explanatory vector and $y_{s,t}$ the observed travel time for segment-direction s at interval t . Linear Regression, Random Forest, and LightGBM estimate $y_{s,t}$ from $x_{s,t}$. For LSTM, the input is the ordered matrix $X_{s,t}^{(L)} = [x_{s,t-L+1}, \dots, x_{s,t}]$, where the selected lookback L is 8, 16, 24, or 32 intervals and the final row corresponds to t . The target sequence $[y_{s,t-L+1}, \dots, y_{s,t-1}]$ is not supplied to the model. This formulation permits the LSTM to learn temporal dependence among explanatory states without using lagged observed or recursively predicted travel times.

6.2. General Modeling Framework

Figure 6 summarizes the transition from the prepared dataset to model evaluation. The ‘concurrent feature set’ box denotes construction of the common base-variable table before architecture-specific input packaging; it does not imply that the LSTM receives only one time row. Detailed source harmonization, quality control, calibration, and feature construction are reported in Sections 3–5; the present section therefore focuses on prediction-unit definition, model configuration, chronological selection, and evaluation.

For every retained segment-direction, the same 31-variable base feature schema and chronological boundaries were applied. Input representation then followed the model family: Linear Regression, Random Forest, and LightGBM consumed the time- t vector, whereas LSTM consumed rolling sequences of those same explanatory vectors. Candidate configurations were fitted on training data, selected exclusively by validation MAE, frozen, and evaluated once on the test period. Segment-direction predictions were then summarized locally and pooled across the study area, followed by statistical comparison and SHAP-

based interpretation. While Figure 6 illustrates this conceptual workflow, the step-by-step computational procedure of the proposed framework is detailed in Algorithm 1.

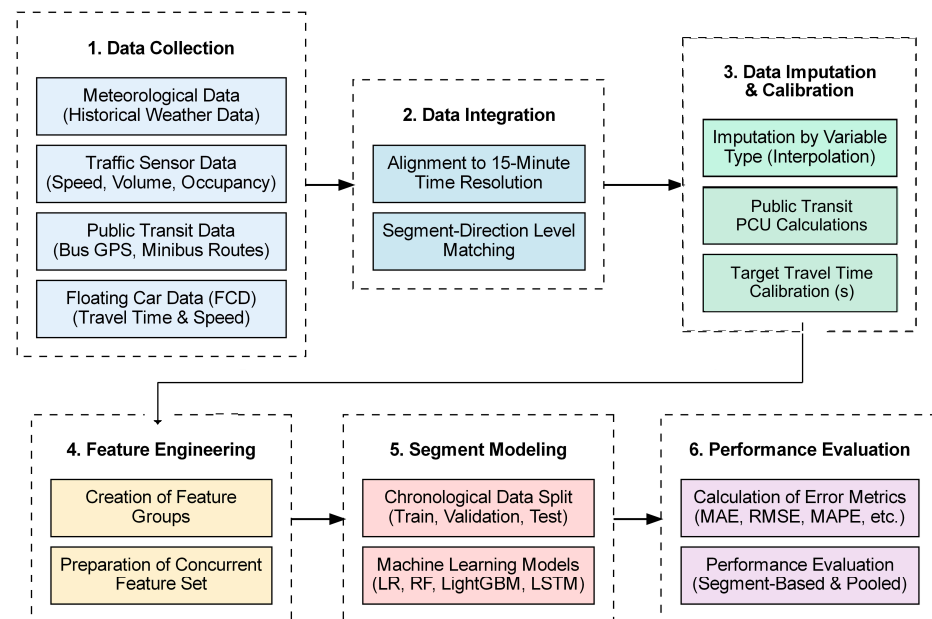


Figure 6. General methodological workflow of this study.

Algorithm 1. Multi-source segment-direction modeling and evaluation framework.

Input: FCD travel times, TCC detector variables, weather records, public transport records, segment-direction map, and chronological split boundaries.

Output: Frozen test predictions, segment-direction metrics, pooled metrics, statistical comparisons, and SHAP summaries.

- 1: Convert source timestamps to common 15 min intervals and map every record to a segment-direction.
 - 2: Apply source-specific quality control, missing-data rules, and verified target calibration.
 - 3: Construct the common 31-variable temporal, traffic-state, meteorological, and PTE/PCU base feature table; exclude target history and engineered lag/rolling columns.
 - 4: For each retained segment-direction s , divide observations chronologically into train, validation, and test periods.
 - 5: Fit preprocessing parameters on train only and apply the frozen transformations to validation and test.
 - 6: Supply time- t feature vectors to LR/RF/LightGBM and rolling base-feature sequences ending at t to LSTM; fit candidates on the train set and select by validation MAE.
 - 7: Freeze the selected configuration; generate test predictions without further tuning.
 - 8: Calculate segment-direction metrics and append predictions to the pooled test table.
 - 9: Compare paired segment-direction errors statistically and calculate SHAP summaries for the selected LightGBM models.
 - 10: Return network-wide accuracy together with spatial and temporal reliability profiles.
-

6.3. Segment-Level Prediction Framework

The final modeling dataset consists of 60 segment-direction data files for which reliable TCC inductive loop detector data could be generated. The same feature groups were used for each segment-direction, but model training and evaluation were conducted independently. Thus, instead of developing a single global model for the entire study area, local models capable of learning the unique traffic dynamics of each segment-direction were created.

This approach avoids the assumption of homogeneous traffic behavior along a corridor and enables direct modeling of spatial differences such as intersection effects, public transport interactions, directional demand differences, and local bottlenecks.

Model performance was evaluated at two levels. First, independent performance metrics were calculated for each segment-direction. Second, the test results of all segment-directions were pooled to evaluate average model performance over the entire study area.

As illustrated in Figure 7, the chronological data split (training, validation, and testing) was applied independently to each segment-direction to ensure rigorous local hyperparameter tuning and isolated performance evaluation.

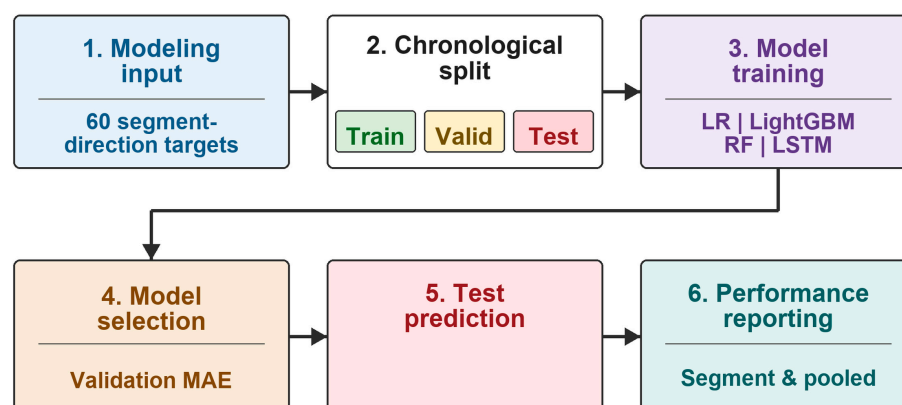


Figure 7. Segment-level modeling and chronological evaluation workflow.

6.4. Prediction Models

Four model families representing different modeling approaches were compared in this study: Linear Regression, Random Forest (RF), Light Gradient Boosting Machine (LightGBM), and Long Short-Term Memory (LSTM).

The model-family selection was designed to jointly evaluate a classical linear baseline, tree-based ensemble methods, and a sequence-based deep learning approach. Linear Regression was used as an interpretable and low-complexity reference model [43]. Random Forest [44] and LightGBM [45,46] represent tree-based ensemble methods capable of modeling nonlinear relationships and variable interactions. LSTM was included as a sequence-based approach and processed rolling windows of the base explanatory vectors, allowing temporal dependence among traffic, temporal, meteorological, and public transport states to be learned without using target history [47].

All models were evaluated with the same chronological split boundaries, target definition, 31 underlying explanatory variables, and performance metrics. The comparison, therefore, controls the information sources and evaluation periods while allowing the input representation required by each model family: a single time- t vector for Linear Regression, Random Forest, and LightGBM, and an ordered lookback-by-feature tensor ending at t for LSTM.

All data processing, modeling, statistical analyses, and visualization were performed in Python 3.11.13 using pandas 3.0.3, NumPy 2.4.6, scikit-learn 1.8.0, LightGBM 4.6.0, TensorFlow 2.21.0, Keras 3.14.1, SciPy 1.17.1, and Matplotlib 3.10.9.

6.5. Hyperparameter and Configuration Selection

Model configurations were determined using only the training and validation datasets, with validation MAE as the selection criterion. For Random Forest and LightGBM, the complete Cartesian product of predefined discrete parameter values was constructed, shuffled using the fixed seed, and a bounded subset was evaluated while always retaining a baseline configuration. This constitutes a reproducible seeded random search over a discrete grid rather than exhaustive grid search or Bayesian optimization. It was selected to

explore interacting parameters at substantially lower computational cost than the full grids while preserving identical candidate-generation logic across segment-directions [30,48].

In the preprocessing stage, missing-value imputation and scaling parameters were learned only from the training data, and the same transformations were applied to the validation and test sets. This prevented data leakage.

For LightGBM, 24 configurations were evaluated. The discrete search space comprised num_leaves {15, 31, 63, 127}, learning_rate {0.01, 0.03, 0.05, 0.08}, n_estimators {600, 1000, 1500, 2200}, min_child_samples {10, 20, 40, 80}, subsample and colsample_bytree {0.70, 0.85, 1.00}, reg_alpha {0, 0.05, 0.20}, reg_lambda {0.10, 0.50, 1.00}, and max_depth {−1, 6, 10}.

For Random Forest, 30 configurations were evaluated from n_estimators {250, 500, 800, 1200}, max_depth {unlimited, 12, 20, 30}, min_samples_split {2, 5, 10}, min_samples_leaf {1, 2, 4}, max_features {square root, log2, 0.8}, and bootstrap {true}.

For LSTM, six prespecified architectures were compared rather than randomly sampled. Lookback windows of 8, 16, 24, and 32 intervals were represented together with 32, 64, or 128 units, one or two LSTM layers, dropout of 0.20 or 0.30, and learning rates of 0.001 or 0.0005. The batch size was 256. A maximum of 80 epochs was used, with early stopping patience of 12; the architecture with the lowest validation MAE was retained separately for each segment-direction.

To ensure reproducibility in all experiments, the seed value was set to 42 for operations involving randomness. This careful configuration process enabled tree-based models to learn complex nonlinear relationships in the high-dimensional feature space [44,46] and enabled the LSTM architecture to learn long-term dependencies in time-series dynamics effectively [47,49].

6.6. Chronological Training, Validation, and Test Design

Because traffic data possess an inherent time-series structure, random data splitting was avoided to prevent temporal information leakage. Instead, the entire dataset was divided chronologically into training, validation, and test sets, as summarized in Table 3.

Table 3. Chronological data split.

Data Partition	Date Range	Obs. Per Seg.-Dir.	Total Obs.	Purpose
Training	5 January 2026 00:30–31 March 2026 23:45	8254	495,240	Learning model parameters
Validation	1 April 2026 00:00–30 April 2026 23:45	2880	172,800	Hyperparameter and configuration selection
Test	1 May 2026 00:00–14 May 2026 23:45	1344	80,640	Final performance reporting
Total	5 January 2026 00:30–14 May 2026 23:45	12,478	748,680	Modeling dataset for 60 segment-directions

This chronological approach ensures that future traffic patterns do not leak into the training process, thereby preserving the temporal integrity of the data and more accurately reflecting real-world operational conditions. In this design, the training set was strictly used to learn model parameters, the validation set for hyperparameter tuning, and the test set exclusively for the final, isolated performance evaluation.

Following this strict chronological separation, the predictive performance of all models on the isolated test set was quantified using established statistical metrics, as detailed in Section 6.7.

6.7. Performance Metrics and Model Comparison

Model performance was evaluated using complementary metrics of absolute error, relative error, explained variance, and systematic bias. Mean absolute error (MAE) was the primary selection metric because it is expressed in seconds and is less sensitive to rare large deviations than RMSE. For n observations, where y_i is the observed travel time, $\hat{y}_i$ is the predicted travel time, and $\bar{y}$ is the mean observed travel time, the metrics are defined as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (8)$$

$$\text{SMAPE} = \frac{100}{n} \sum_{i=1}^n \frac{2|y_i - \hat{y}_i|}{|y_i| + |\hat{y}_i|} \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (10)$$

$$\text{ME} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \quad (11)$$

MAE and RMSE quantify typical and squared-error-sensitive magnitudes, respectively. MAPE and SMAPE express relative error as percentages; all observed travel times were positive, and SMAPE additionally normalizes by both observed and predicted magnitudes. R^2 measures variance explained relative to the observed mean. Mean error (ME) follows the observed-minus-predicted convention of the analysis, so positive values indicate underprediction and negative values indicate overprediction [48,49].

Reporting all six metrics prevents conclusions from depending on a single loss function: MAE was used for model selection, RMSE highlighted tail-sensitive failures, percentage errors supported scale-aware interpretation, R^2 summarized explanatory power, and ME identified directional bias. To ensure comprehensive evaluation, these metrics were computed at two aggregation levels: independently for each segment-direction to examine spatial performance variability, and across the pooled test set to determine the overall model rankings for the entire study area.

7. Results

This section first presents the overall model performance obtained by pooling the test observations of all segment-directions. The results are then evaluated in detail at the segment-direction, corridor, and intra-day period levels to examine the spatial and temporal variability of model performance.

7.1. Overall Model Performance

The final test results show that tree-based ensemble models are the strongest model family in terms of overall performance. LightGBM yielded the lowest MAE on the pooled test set; however, the difference between LightGBM and Random Forest was only 0.002 s. Therefore, the results should be interpreted not as the clear dominance of a single model, but as evidence that tree-based models perform consistently well for the present data structure.

The LSTM model produced lower errors than Linear Regression but lagged behind LightGBM and Random Forest in pooled test performance. LightGBM and Random Forest produced approximately 17.6% lower MAE than Linear Regression, while LSTM's

improvement was approximately 15.8%. The low mean error values relative to the absolute error levels indicate that the models do not have strong systematic bias; nevertheless, the LightGBM and Random Forest models showed a small tendency toward underprediction.

Table 4 compares the four model families on the same number of observations and the same chronological test period. MAE and RMSE values show that tree-based models reduce error magnitude, while R^2 values indicate that these models explain the variability in the target variable better than Linear Regression. However, because the difference between LightGBM and Random Forest is minimal, the overall performance interpretation should be based on behavioral differences between model families rather than on a single winning model.

Table 4. Overall model performance on the final test set.

Model	Test Size (<i>n</i>)	MAE (s)	RMSE (s)	MAPE (%)	SMAPE (%)	R^2	Mean Error (s)
LightGBM	80,640	6.750	17.147	8.213	8.036	0.790	0.285
Random Forest	80,640	6.752	17.176	8.229	8.047	0.790	0.246
LSTM	80,640	6.898	17.456	8.403	8.243	0.783	0.109
Linear Regression	80,640	8.189	18.345	10.162	9.963	0.760	−0.115

Because the pooled errors of LightGBM and Random Forest were almost identical, computational evidence was also examined. Existing tuning logs show median per-segment search times of 2.896 s for 24 LightGBM candidates and 19.930 s for 30 Random Forest candidates; the six-configuration LSTM search required 116.374 s. Thus, in the recorded experimental environment, the Random Forest search was 6.9 times and the LSTM search 40.2 times slower than the LightGBM search. These values cover candidate fitting and validation and are not production inference benchmarks. Prediction latency was not separately logged; however, real-time deployment must additionally account for the number and depth of fitted trees, sequential LSTM evaluation, model storage, and the need to maintain 60 local models. When accuracy is statistically indistinguishable, the markedly lower recorded tuning cost gives LightGBM a practical retraining advantage, whereas a deployment-specific latency benchmark remains necessary before operational selection.

The metrics presented in Table 4 show the overall performance obtained by pooling all segment-direction test observations. However, these results alone do not reveal how error levels are distributed within the study area. Therefore, the overall performance evaluation should be interpreted together with Figure 8, which shows segment-direction-level error distributions.

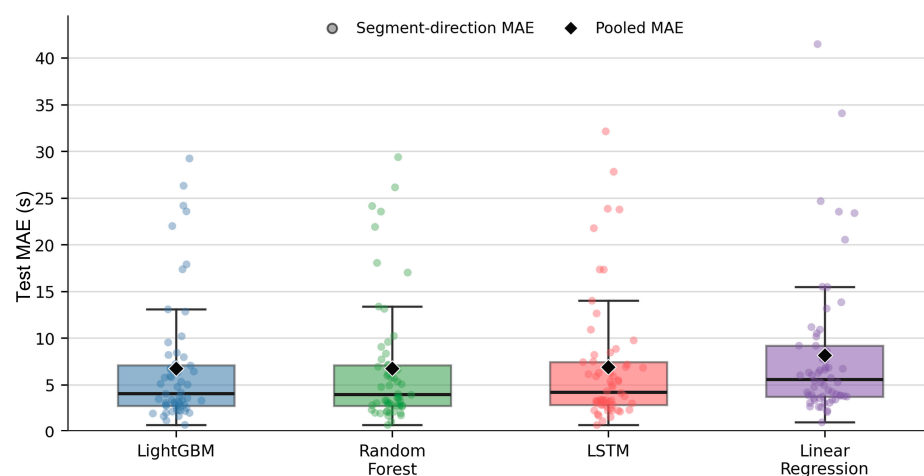


Figure 8. Distribution of segment-direction-level test MAE by model family.

In Figure 8, colored points represent the test MAE values of the 60 segment-directions for each model family, boxes represent the distribution of these values, and black diamonds

represent pooled test MAE values. The distributions of LightGBM and Random Forest are very close to each other, whereas Linear Regression has a higher median error level and wider error spread. This finding supports the overall superiority of tree-based models but also shows clear performance differences among segment-directions. Therefore, the next subsection details model performance at the segment-direction level.

7.2. Segment-Direction-Level Model Comparison

Although pooled test results are useful for summarizing overall performance, segment-direction-level results show that model performance is not homogeneous throughout the study area.

Figure 9 shows the distribution of the model family that produced the lowest MAE on the test set for each segment-direction.

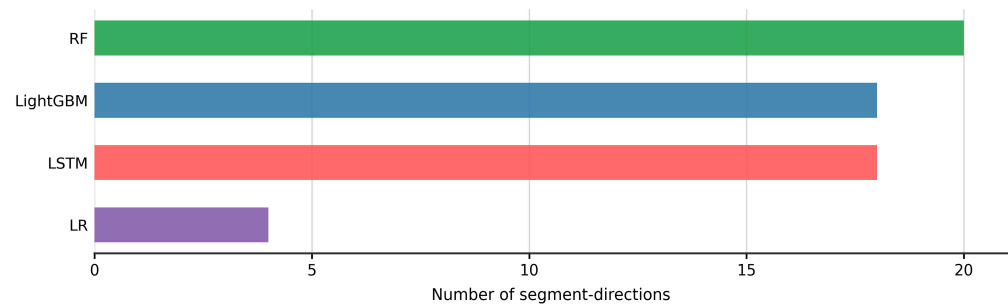


Figure 9. Distribution of the model yielding the lowest test MAE by segment-direction.

As seen in the figure, although LightGBM produced the lowest MAE on the pooled test set, Random Forest achieved the largest number of local best results at the segment-direction level. Random Forest produced the lowest test MAE in 20 segment-directions, LightGBM and LSTM in 18 segment-directions each, and Linear Regression in 4 segment-directions.

This distribution shows that the overall performance ranking and the segment-direction-level best-model distribution do not provide the same information. The fact that LSTM produced the lowest error in 18 segment-directions despite ranking behind tree-based models overall indicates that sequence-based model structures can capture local patterns in some segment-directions. Conversely, the limited number of segment-directions in which Linear Regression stands out suggests that nonlinear model structures are more suitable for most of the study area.

The existing LSTM configuration search provides further guidance on where sequence modeling was useful. Lookbacks of 8, 16, 24, and 32 intervals were selected for 4, 35, 14, and 7 segment-directions, respectively. Across the six fixed architectures, pooled test MAE varied only from 6.957 to 7.033 s; segment-specific validation selection reduced it to 6.898 s. Segment-directions for which LSTM was locally best had slightly higher median lag-1 travel time autocorrelation (0.895 versus 0.862), mean travel time (75.45 versus 69.47 s), and travel time standard deviation (13.98 versus 12.26 s) than the remaining links. These descriptive differences suggest that LSTM can be locally useful under more persistent yet variable regimes, but their modest magnitude also explains why no single lookback or sequence architecture dominated the network.

The segment-direction-level error distribution also supports this heterogeneity. The median segment-direction MAE is 3.971 s for Random Forest, 4.083 s for LightGBM, 4.222 s for LSTM, and 5.589 s for Linear Regression. The corresponding mean MAE values are 6.752, 6.750, 6.898, and 8.189 s. The fact that mean values are markedly higher than median values indicates that overall performance is pushed upward by a limited number of high-error segment-directions.

The complete list of the models yielding the lowest test MAE for each segment-direction and the corresponding error metrics is presented in Appendix A Table A1 to avoid extending the main text.

The segment-direction-level findings show that model selection should not only be based on pooled test MAE. Although LightGBM and Random Forest are strong candidates for general application, segment-level evaluation reveals that different model families can be more successful on specific road sections. This result motivates the corridor-level examination of error patterns in the next subsection.

7.3. Corridor-Level Performance Patterns

The performance of the models was further evaluated at the corridor level. When segment-direction-level results are aggregated by corridor, the model errors are distributed heterogeneously throughout the study area. Table 5 presents data from the model that produced the lowest MAE on the test set for each segment-direction. The table should not be used to compare the four model families, but to show how final segment-direction results differ at the corridor scale. Because the complete segment-direction list is provided in Appendix A Table A1, only the corridor-level summary and distinguishing patterns are presented in the main text.

Table 5. Corridor-based summary of segment-direction test performance.

Corridor	Seg.-Dir. <i>n</i>	Leading Model	Mean MAE (s)	Median MAE (s)	Max. MAE (s)	R^2
Corridor 1	13	LGBM/RF (5 each)	5.334	2.999	17.325	0.581
Corridor 2	8	LGBM/LSTM (3 each)	8.147	5.693	29.234	0.205
Corridor 3	12	LSTM/RF (4 each)	5.183	4.633	12.640	0.450
Corridor 4	14	LGBM (5)	7.114	2.878	23.852	0.375
Corridor 5	5	LSTM (3)	7.453	2.841	26.141	0.542
Corridor 6	8	LGBM/RF (3 each)	7.714	6.438	21.768	0.724

Two distinct error patterns emerge from the results presented in Table 5. Corridor 2 appears to be among the most challenging corridors to model, as evidenced by both its relatively high median MAE and its low average R^2 value. In contrast, although Corridor 6 exhibits a higher median MAE, its average R^2 reaches 0.724, indicating that while travel time variability is greater in this corridor, the models are still able to capture the underlying variability patterns with reasonable accuracy. For Corridors 4 and 5, median error values remain comparatively low; however, the substantially higher maximum MAE values suggest the presence of localized prediction failures. Consequently, the reduction in overall model performance within these corridors is not attributable to corridor-wide deficiencies, but rather to concentrated error clusters associated with specific segment-direction combinations.

A targeted audit of the already-prepared modeling tables found no residual missing detector cells in Corridor 2. Its median detector repeat rate (0.005) and zero rate (0.000) were also comparable to the other corridors (0.006 and 0.001), so processed-table completeness alone does not explain the low R^2 . Corridor 2 nevertheless had higher median target variability (coefficient of variation 0.264 versus 0.176) and lag-1 autocorrelation (0.894 versus 0.862), while its median segment-direction R^2 was only 0.134 compared with 0.617 elsewhere. The most difficult link, R_H_Kosk–Tiyatro, combined the highest target variability (coefficient of variation 0.474) with a weaker maximum absolute detector-target Spearman association (0.574). This pattern points to travel time dynamics not fully represented by the concurrent detector variables or to unobserved local operations rather than to residual missingness in the processed table. Raw detector calibration and failure logs were not

available, so a sensor-calibration contribution cannot be ruled out and should be checked before deployment.

The highest-error segment-directions were identified as R_H_Kosk–Tiyatro, R_Adliye–Aydem, G_Hal–Toki, R_Hallaclar–ASM, and G_YSE–Incilipinar. The lowest test MAE values in these segment-directions range from 21.768 to 29.234 s. These segments should be interpreted as candidates for engineering diagnosis rather than as simple model failures. Around such links, error growth may be associated with signal timing variability, queue spillback near intersections, stop and boarding activity, turning movements, or unobserved incidents and road works [50]. Because detailed signal plan and incident records were not available in the modeling table, these explanations cannot be tested causally here; nevertheless, they identify the local conditions that should be checked before operational deployment.

Figure 10 presents the corridor-level heterogeneity summarized numerically in Table 5 along the segment-direction order. Rows represent corridor-direction pairs, and columns represent the segment order within each corridor.

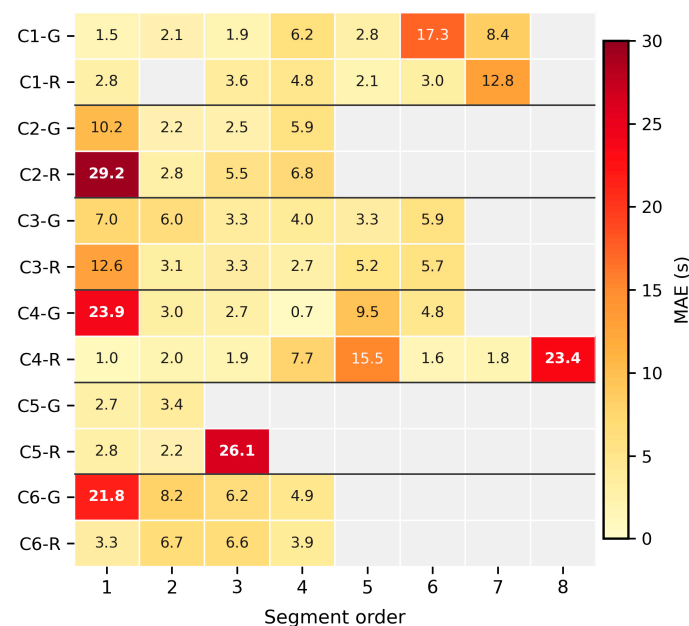


Figure 10. Test MAE error map by corridor and segment-direction order. Cell values represent the lowest test MAE for each modeled segment-direction; darker colors indicate larger errors. Gray cells denote positions without a physical segment or outside the final modeling scope.

This visualization provides a complementary spatial error summary to Table 5, revealing that high error values are not evenly distributed but rather concentrated in specific corridor-direction and segment-order positions.

7.4. Prediction Behavior on a Representative Segment-Direction

Aggregated performance metrics reveal general differences among models but do not show how predictions behave over time. Therefore, a representative segment-direction was examined to evaluate the temporal agreement between observed and predicted travel time series. The G_Yuksektepe–Bagbasi segment-direction was selected because its LightGBM test MAE (4.150 s) is close to the median LightGBM MAE for segment-directions (4.083 s).

This choice provides a balanced visual evaluation in two respects. First, the selected segment-direction is neither an extremely low-error nor a high-error case. Second, although Random Forest produced the lowest MAE locally on this segment, the difference between Random Forest and LightGBM was only 0.128 s. Therefore, using LightGBM in Figure 11, the strongest model in the pooled test results, is consistent with the overall

model comparison and is not misleading for the selected segment-direction. For the same segment-direction, the test MAE values of Random Forest, LSTM, and Linear Regression were 4.022, 4.123, and 4.743 s, respectively.

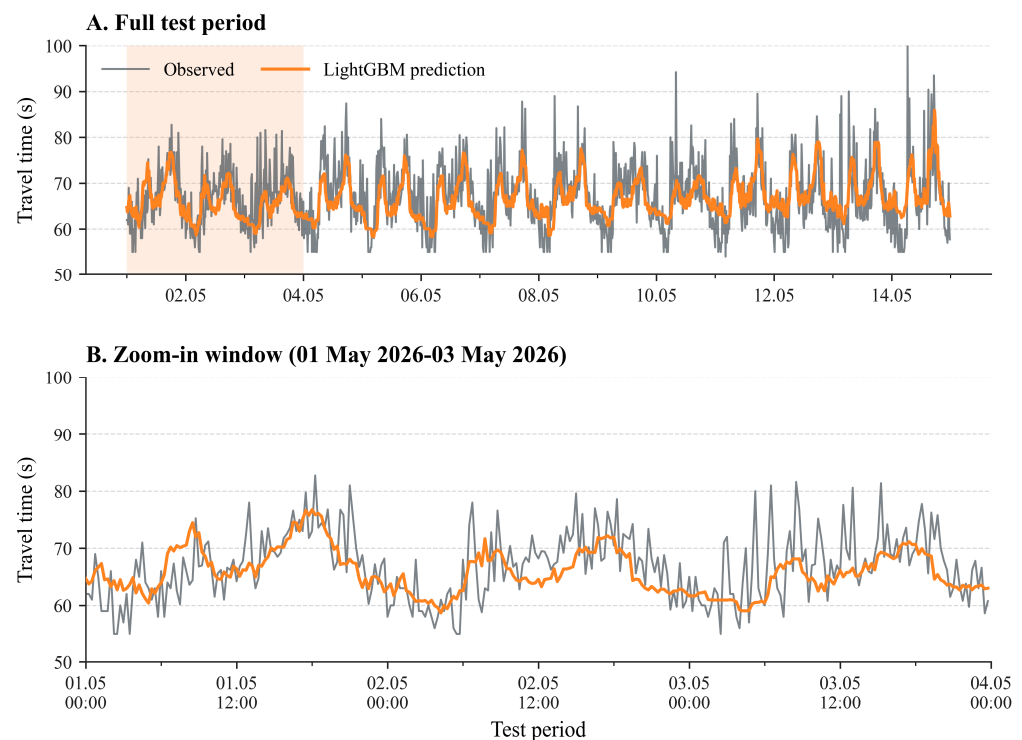


Figure 11. Observed travel time and LightGBM prediction during the test period for the representative G_Yuksektepe–Bagbasi segment-direction. The light shaded region in (A) indicates the interval enlarged in (B).

Figure 11 presents the observed travel time and LightGBM prediction in a two-panel structure. Panel A shows the full test period, preserving the overall temporal context and allowing the long-range stability of the prediction series to be assessed. Panel B provides a zoomed view of a three-day interval from 1 May to 3 May 2026, covering a weekday-to-weekend transition. The shaded region in Panel A indicates the time interval enlarged in Panel B. This design enables the evaluation of both the general tracking behavior over the full test period and the short-term response of the model to daily fluctuations in the same figure. This visualization is not intended to generalize from a single segment; rather, it shows how the numerical performance summarized in Table 4 and Figure 8 appears at the time-series level. In the selected segment-direction, the mean observed travel time is 67.26 s and the mean LightGBM prediction is 66.63 s. The median absolute error is 3.32 s. In 75% of the test observations, the absolute error remains below 5.70 s, and the maximum absolute error is 30.08 s.

The purpose of Figure 11 is to assess whether a representative low-MAE model follows both recurrent levels and short-duration deviations. The two-panel visualization shows that LightGBM consistently follows the general travel time level and daily pattern, including morning and evening increases, but partially smooths sudden spikes. The highest observed value is 100 s, whereas the maximum prediction is 85.90 s. Such smoothing may be acceptable for routine information provision when average reliability is the priority, but it is not sufficient for risk-sensitive uses such as emergency routing or automated incident response. Operational deployment should therefore complement point predictions with calibrated uncertainty estimates; the cited multi-horizon study illustrates one possible

approach through quantile regression [31]. Cases with high predictive uncertainty should then be flagged for operator review.

7.5. Intra-Day Error Patterns

In addition to spatial heterogeneity, model reliability also differs clearly across intra-day traffic operation periods. This subsection examines temporal error behavior using the LightGBM model, which produced the lowest MAE in the pooled test results. Intra-day periods were categorized as night, morning peak, daytime, evening peak, and evening according to hour ranges, and absolute error values were calculated for each test observation.

Figure 12 depicts a summary of the LightGBM test absolute error distribution by intra-day period. In the visualization, thick horizontal lines represent the middle 50% error range of the observations, circles indicate medians, diamonds denote means, and red vertical marks show the 90th percentile error threshold.

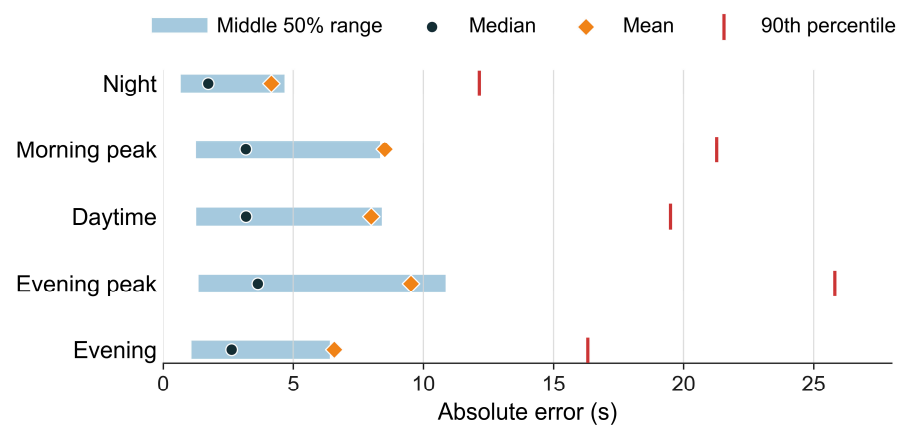


Figure 12. Summary of LightGBM test absolute error distribution by intra-day period.

The mean absolute error is lowest during the night period at 4.158 s. In contrast, the error rises to 8.515 s in the morning peak and reaches its maximum at 9.523 s in the evening peak. The mean error is 7.993 s during the daytime period and 6.575 s during the evening period. Median and upper-tail values clarify why average error levels increase during peak hours. The median absolute error is 1.724 s at night and rises to 3.636 s in the evening peak. More importantly, while 90% of test observations remain below an absolute error of 12.150 s during the night period, this threshold widens to 21.267 s in the morning peak and 25.814 s in the evening peak.

Hourly evaluations further support this pattern: the lowest hourly mean absolute error was 3.55 s at 04:00, whereas the highest was 10.63 s at 15:00, with 08:00 and 17:00 also standing out with elevated error levels.

These findings demonstrate that the challenge during peak periods is not only a general increase in typical error levels but also the emergence of less frequent yet larger deviations. Consequently, model performance is highly sensitive not only to spatial segment locations but also to the temporal variability of traffic demand throughout the day.

7.6. SHAP-Based Interpretation of Feature Group Contributions

In addition to overall performance and error-pattern analyses, the information groups used by the LightGBM model, which produced the lowest MAE on the pooled test set, were examined. For each segment-direction, the model was retrained using the LightGBM hyperparameters selected on the validation set, and SHAP (Shapley Additive Explanations) contributions were calculated for the test observations. Individual variable contributions were

grouped under traffic state, public transport exposure, temporal variables, and meteorology; within each segment-direction, they were converted to relative contribution percentages.

Figure 13 shows the SHAP-based relative absolute contribution of feature groups in the LightGBM model. The SHAP approach is widely used, especially for tree-based models, because it allows local contributions to be aggregated into more general interpretations [51,52].



Figure 13. Distribution of relative absolute SHAP contributions across 60 segment-directions by feature group. Points represent individual segment-direction values; diamonds and thick horizontal lines indicate the medians and interquartile ranges, respectively.

In Figure 13, points show individual segment-direction values, diamonds show the group medians, and thick horizontal lines show the interquartile ranges. Temporal variables rank first with a median contribution of 49.5%, while the median contribution of traffic-state variables is 36.2%. The median contributions of meteorology and public transport exposure groups are 7.2% and 3.3%, respectively. This result should be interpreted carefully: a high temporal contribution means that the model learns the recurrent rhythm of urban traffic, not that real-time traffic state is unimportant. The 36.2% contribution of traffic-state variables indicates that loop detector measurements provide the main adjustment around that rhythm, while meteorology and public transport variables add smaller but complementary signals. To examine the model's decision mechanism in greater detail, the top 10 SHAP contributions at the individual-variable level are presented in Figure 14.

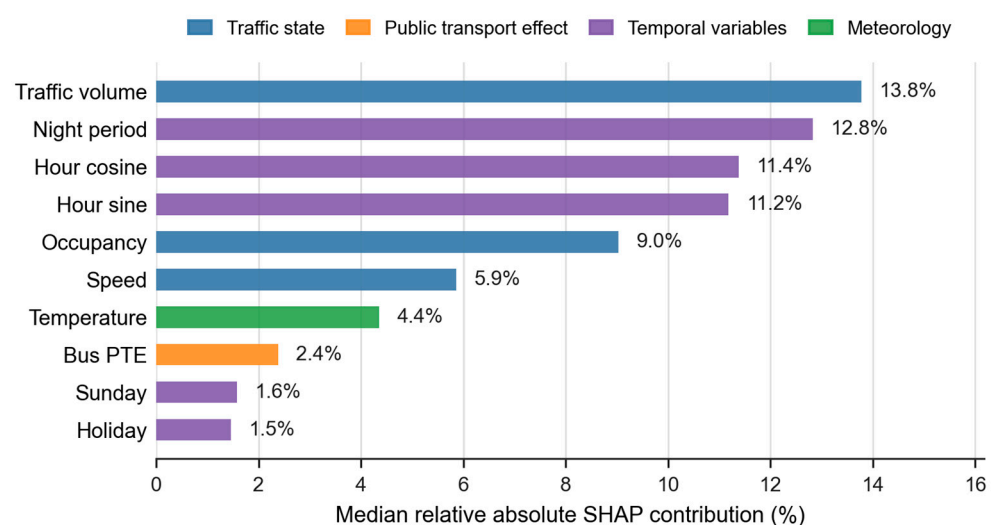


Figure 14. SHAP importance ranking of the top 10 variables with the highest mean effect on LightGBM predictions.

At the individual-variable level, traffic volume stands out as the most influential feature, with a median contribution of 13.8%. It is followed by the night period with 12.8%. The high contribution of the night-period indicator shows that the model captures recurrent transitions at the beginning of daily mobility, including the start of minibus services at 06:00 in the urban transport network. In addition, the clear effects of hour cosine (11.4%) and hour sine (11.2%), obtained through trigonometric transformations, confirm that cyclic encoding contributes to model performance by preserving the periodic nature of time. The Bus PTE feature, which appears among the top 10 variables with a median contribution of 2.4%, supports the validity of representing public transport exposure as a traffic-load effect. These SHAP contributions provide model-internal explanations and should not be interpreted as direct causal effects.

To complement the global SHAP ranking with a local response-shape analysis, Figure 15 reports 1344 chronological test observations for the representative G_Yuksektepe–Bagbasi segment-direction used in Figure 11. The selected LightGBM model was refitted with its validation-selected configuration and reproduced the reported test MAE (4.150 s). The contribution of traffic volume remained negative or close to zero across low and moderate values but rose sharply at the upper end of the recorded volume range. Occupancy showed a related nonlinear transition from negative contributions at low values to positive contributions at intermediate and high values. Coloring each panel by the paired traffic-state variable shows that these responses co-occur with changing occupancy-volume regimes. The plots describe the fitted model's conditional associations and threshold behavior; they do not establish causal traffic-flow effects.

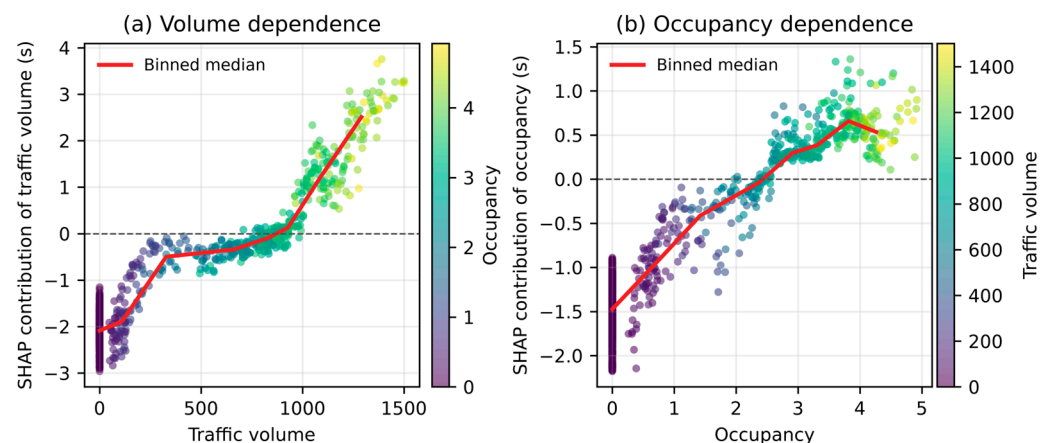


Figure 15. SHAP dependence of traffic volume and occupancy for the representative G_Yuksektepe–Bagbasi segment-direction. Points represent 1344 chronological test observations, color denotes the paired traffic-state variable, and red lines show binned medians.

A leave-one-group-out ablation was then performed for all 60 segment-directions. For each segment-direction, the full LightGBM model and four reduced models were refitted after separately removing the temporal, traffic-state, meteorological, or public transport exposure group. The previously selected segment-specific hyperparameters were held fixed to isolate feature removal; median imputation was fitted on training data only, validation data controlled early stopping, and the chronological test set was used once for reporting. The refitted full specification exactly reproduced the pooled MAE of 6.750 s. Table 6 reports the accuracy change caused by each group removal together with paired segment-level Wilcoxon tests and Holm correction.

Table 6. Leave-one-group-out LightGBM ablation on the chronological test set.

Removed Group	Pooled MAE (s)	Delta MAE (s)	Median Segment Delta MAE (s)	Segments with Higher MAE	Holm p
Temporal	7.891	+1.141	+0.949	58/60	<0.001
Traffic state	6.942	+0.192	+0.179	42/60	<0.001
Meteorology	6.866	+0.116	+0.061	36/60	0.006
Public transport	6.811	+0.061	+0.029	42/60	0.001

The ablation results support, but also refine, the global SHAP interpretation. Removing temporal variables produced the largest pooled degradation (+1.141 s MAE) and worsened 58 of 60 segment-directions. Traffic-state removal increased pooled MAE by 0.192 s. Meteorology and public transport produced smaller changes, but their removal still increased pooled MAE by 0.116 and 0.061 s, respectively; public transport removal worsened 42 segment-directions despite its 3.3% median SHAP attribution. All four paired differences remained significant after Holm correction (adjusted $p \leq 0.006$). Thus, temporal recurrence supplies the strongest marginal information, detector variables provide the main real-time adjustment, and meteorology and public transport add smaller but statistically consistent complementary value.

7.7. Statistical Evaluation of Differences Among Models

Nonparametric comparison tests were applied to test MAE values to evaluate whether differences in model performance were statistically significant at the segment-direction level. The error values of the four models on the same 60 segment-directions were treated as paired measurements. In the first stage, the Friedman test was used; then, Wilcoxon signed-rank tests were applied for pairwise comparisons. To control for multiple comparisons, the p -values of the Wilcoxon tests were adjusted using the Holm method. This test selection enables nonparametric evaluation of multiple models compared on the same data partitions and controls the effect of multiple testing [53].

The Friedman test showed an overall significant difference among the segment-direction-level MAE distributions of the four models ($\chi^2(3) = 80.94$; $p < 0.001$). Pairwise comparisons revealed that LightGBM, Random Forest, and LSTM produced statistically significantly lower MAE values than Linear Regression (Holm-adjusted $p < 0.001$). The differences between LSTM and Random Forest ($p = 0.061$) and between LSTM and LightGBM ($p = 0.090$) were not statistically significant at the 0.05 level. Similarly, the difference between LightGBM and Random Forest was not statistically significant (Holm-adjusted $p = 0.551$). This result demonstrates that although LightGBM produced the lowest MAE on the pooled test set, its performance advantage over Random Forest was not statistically distinct. The findings, therefore, point to the strong and stable performance of the tree-based model family for this data structure rather than to the absolute superiority of a single model.

7.8. General Evaluation of the Findings

The comprehensive evaluation demonstrates that model performance cannot be fully captured by a single aggregated error metric. While Table 4 and Figure 8 summarize overall model performance, Figure 9 highlights the localized best-model distributions at the segment-direction level. Statistical comparisons (Section 7.7) confirm that the overall difference between LightGBM and Random Forest is not significant. Furthermore, Table 5 and Figure 10 illustrate that local performance differences are distributed heterogeneously across corridors, whereas Figures 11 and 12 project this heterogeneity onto the temporal axis. Finally, Figures 13–15 and Table 6 unpack the model's internal logic, detailing global

feature importance, local nonlinear responses, and the marginal accuracy contributions of specific feature groups.

Synthesizing these multi-level evaluations reveals several key insights. First, although LightGBM and Random Forest yielded the lowest errors on the pooled test set, their performance difference was not statistically significant, as detailed in Section 7.7. Second, the spatial distribution of the best-performing models diverged from the overall ranking: Random Forest produced the lowest test MAE in 20 segment-directions, while LSTM and LightGBM did so in 18 segment-directions each. Third, prediction errors were not random but concentrated within specific corridor-directions and segment orders. Fourth, temporal degradation was evident, with model errors increasing particularly during morning and evening peak periods. Fifth, explainability analyses confirmed that temporal and traffic-state variables made the dominant contributions to LightGBM predictions.

Ultimately, these findings establish tree-based ensemble models as robust candidates for the study area, while underscoring that predictive performance is deeply intertwined with spatial (segment-direction and corridor) and temporal (intra-day) traffic characteristics. Therefore, the results should be interpreted through the lens of the underlying data structure and local traffic dynamics rather than solely through algorithm selection. The next section discusses these findings within the broader context of the literature, transportation engineering principles, and ITS.

8. Discussion

The principal finding of this study is that while the tabular multi-source structure naturally favors tree-based ensemble models, no single algorithm is universally superior across an entire urban network. This finding is consistent with studies emphasizing the importance of nonlinear interactions among traffic state, temporal, weather, and public transport variables in travel time prediction [40,54]. The strong overall performance of LightGBM and Random Forest aligns with their established flexibility in capturing threshold effects and segment-specific nonlinear patterns [44,46]. Conversely, the lower pooled performance of LSTM in this study should not be interpreted as evidence that sequence-based models are generally weak. Rather, the specific segment-direction sample size, the short chronological test window, and the strong explanatory signal in tabular variables favored tree-based architectures in the present setting. Therefore, the core contribution is not merely the selection of LightGBM, but the demonstration that algorithm choice must be critically evaluated alongside segment-level reliability.

The integration of FCD-based observations, inductive loop data, meteorological conditions, and public transport indicators at a common 15 min resolution demonstrates that fragmented institutional data in medium-sized cities can be unified into a robust analytical framework. Global SHAP and leave-one-group-out ablation provide complementary evidence: the largest accuracy loss occurred after temporal variable removal (+1.141 s pooled MAE), followed by traffic state (+0.192 s), meteorology (+0.116 s), and public transport (+0.061 s). The smaller groups were not redundant; meteorology and public transport removals remained significant after Holm correction, and public transport removal worsened 42 of 60 segment-directions. These results describe predictive information within the fitted data structure rather than causal effects, but they confirm that recurrent temporal structure and real-time detector states dominate, while weather and public transport provide smaller, consistent adjustments.

Evaluating model performance through a segment-direction approach fundamentally refines corridor-level traffic modeling. Traditional corridor models often report acceptable average errors while masking localized unreliability [55]. In contrast, this approach identified that while Corridor 2 was globally challenging, errors in Corridors 4 and 5 were highly

localized to specific segments despite lower median corridor errors. For Traffic Control Centers (TCCs), this spatial and temporal heterogeneity translates directly into operational strategy. Operational decisions are made on specific links and time intervals rather than network-wide averages [56]. In low-error segments, predictions can be integrated directly into decision-support systems for variable message signs, corridor monitoring, or signal coordination. Conversely, predictions for high-error segments should function as advisory flags, prompting operators to investigate localized physical constraints, queue spillbacks, road works, or unusual events before initiating automated responses.

The asymmetric observation of buses and minibuses is a further measurement limitation. Bus passage and speed information was obtained from vehicle GPS records, whereas minibus passages were schedule-based and minibus speed was assumed to be 80% of contemporaneous segment FCD speed with a 5-km/h floor. This provides a consistent segment-time proxy but may misrepresent realized minibus delays, missed or additional services, stop dwell time, and lane-blocking effects. The minibus PCU signal should therefore be interpreted as scheduled exposure rather than direct vehicle observation, and future work should validate it using minibus GPS and stop-level dwell records.

The relatively low median PTE contribution (3.3%) should not be interpreted as evidence that public transport is operationally irrelevant. PTE shares information with recurring temporal schedules and detector-observed traffic states, while the current PCU-based representation cannot directly capture stop density, dwell time, or localized boarding and alighting queues. These factors may shift part of the public transport signal to correlated temporal or occupancy variables. Future data collection should therefore incorporate stop-level dwell time, passenger exchange, stop density, and lane-blocking indicators to represent the nonlinear local effects of buses and minibuses more directly.

A practical TCC implementation could expose one record per segment-direction and 15 min interval through an application programming interface, containing the point prediction, model identifier, recent validation/error profile, and a reliability or uncertainty flag. A dashboard could then display low-risk predictions directly, use warning colors for historically high-error links or peak periods, and require operator confirmation before adaptive signal or traveler-information actions when uncertainty exceeds a locally calibrated threshold. Confidence thresholds should be calibrated from prediction intervals and operational loss functions rather than assigned as universal fixed values.

Despite its operational utility, this framework has several limitations. The modeling period (5 January to 14 May 2026) captures dense short-term dynamics but not a full annual cycle, summer demand patterns, or severe winter conditions; full-year external validation is needed before seasonal generalizability can be claimed. Hourly weather observations were repeated across four 15 min intervals, so within-hour precipitation onset or cessation could not be represented, and the incremental meteorological contribution may be attenuated. The final scope was restricted to segment-directions with reliable TCC explanatory variables, improving internal data quality but limiting network coverage. Explicit crash, closure, road work, special event, signal plan, lane configuration, and stop density variables were unavailable, preventing causal diagnosis of isolated high-error links. Finally, independent segment-direction models isolate local behavior but do not represent queue spillback, coordinated signals, or other dependencies among neighboring links.

Scaling the present local-model architecture to several hundred segment-directions would increase model storage, tuning, monitoring, and periodic retraining approximately in proportion to the number of retained links, although segment-level fitting is naturally parallelizable. Future evaluations should compare this design with shared multi-task models and spatial-temporal architectures, including Temporal Convolutional Networks, graph neural networks, Graph Attention Networks, and Temporal Fusion Transformers.

Recent work on trajectory granularity optimization also shows that spatial representation choices can affect both prediction accuracy and computational efficiency in urban travel time prediction [33]. Such alternatives should therefore be assessed against tree ensembles using chronological accuracy, calibration, inference latency, memory, and maintainability rather than predictive error alone.

9. Conclusions

This study developed and evaluated a multi-source machine learning framework for short-term travel time prediction at a 15 min segment-direction resolution in the Denizli city center. By integrating FCD travel times, TCC detector variables, meteorological variables, and public transport indicators into a unified tabular structure, robust analyses were performed on 60 segment-directions with reliable explanatory variables. The results indicate that while tree-based ensemble models are highly competitive for this multi-source tabular prediction problem, segment-direction-level evaluations caution against adopting a single universal model. Specifically, Random Forest produced the best local result in 20 segment-directions, LSTM in 18, LightGBM in 18, and Linear Regression in 4. This distribution demonstrates that model performance is highly sensitive to local traffic conditions and data characteristics. Furthermore, corridor and intra-day analyses revealed that prediction error is spatially and temporally heterogeneous, with uncertainty concentrated in specific links and peak operating periods.

The primary contribution of this study is a chronologically evaluated and operationally interpretable prediction framework tailored for medium-sized cities. It equips TCC operators with a dual-layered tool: accurate short-term travel time predictions and a localized reliability profile indicating where automated interventions are appropriate versus where manual operator review is essential. Future studies should extend this framework by incorporating long-term seasonal data, labeled traffic incidents, geometric variables, prediction-interval evaluation, and spatially aware modeling structures. These extensions will enhance both the scientific explanatory power and the practical usability of segment-direction travel time prediction models.

Beyond predictive performance, the framework has an enabling role in sustainable urban mobility. More reliable short-term information can support dynamic route guidance, traffic management, and public transport operations, which may in turn reduce avoidable delay, fuel use, and emissions. However, the present study evaluates prediction accuracy rather than a deployed control intervention; it therefore does not quantify fuel or emission reductions and should not be interpreted as direct evidence of an achieved environmental benefit. Future field or simulation experiments should connect prediction-informed control actions to measured delay, fuel consumption, and emission outcomes.

Author Contributions: Conceptualization, M.E.K. and Y.S.M.; Methodology, M.E.K. and Y.S.M.; Software, M.E.K.; Validation, M.E.K. and Y.S.M.; Formal Analysis, M.E.K.; Investigation, M.E.K. and Y.S.M.; Resources, M.E.K. and Y.S.M.; Data Curation, M.E.K.; Writing—Original Draft Preparation, M.E.K.; Writing—Review and Editing, Y.S.M.; Visualization, M.E.K.; Supervision, Y.S.M.; Project Administration, Y.S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Scientific Research Coordination Unit of Pamukkale University under project number 2026FE012. The researchers express their gratitude to Pamukkale University Scientific Research Projects Coordination Unit for its support.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

Abbreviations

The following abbreviations are used in this manuscript:

API	Application programming interface
ARIMA	Autoregressive integrated moving average
AVL	Automatic vehicle location
Corr.	Corridor
FCD	Floating car data
GPS	Global Positioning System
ITS	Intelligent Transportation Systems
LGBM	Light Gradient Boosting Machine
LR	Linear Regression
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
Obs.	Observations
PCU	Passenger Car Unit
PTE	Public Transport Effect
R^2	Coefficient of determination
RF	Random Forest
RMSE	Root Mean Square Error
Seg.	Segment
Seg.-dir.	Segment-direction
SHAP	SHapley Additive exPlanations
SMAPE	Symmetric Mean Absolute Percentage Error
TCC	Traffic Control Center
XGBoost	Extreme Gradient Boosting

Appendix A

Table A1. Segment-direction best-model test performance.

Seg.-Dir.	Corr.	Seg. No.	Best Model	MAE (s)	RMSE (s)	R^2	ME (s)
R_Adliye_Loj-Adliye	1	1	RF	2.81	6.15	0.85	2.15
G_Adliye-Adliye_Loj	1	1	LSTM	1.54	5.04	0.60	0.79
G_Adliye_Loj-Sirinkoy	1	2	RF	2.11	11.09	0.50	0.68
R_M.Ormani-Sirinkoy	1	3	LGBM	3.56	7.04	0.85	2.13
G_Sirinkoy-M.Ormani	1	3	LGBM	1.92	7.41	0.52	0.35
R_Yenisehir-M.Ormani	1	4	LSTM	4.77	14.23	0.63	2.25
G_M.Ormani-Yenisehir	1	4	RF	6.18	20.30	0.64	4.19
R_Belkon-Yenisehir	1	5	RF	2.09	12.64	0.45	1.64
G_Yenisehir-Belkon	1	5	LGBM	2.77	7.32	0.68	0.34
R_Basak-Belkon	1	6	RF	3.00	4.48	0.83	2.11
G_Belkon-Basak	1	6	LSTM	17.32	31.95	0.69	3.33
R_Tiyatro-Basak	1	7	LGBM	12.84	15.49	−0.07	−8.62
G_Basak-Tiyatro	1	7	LGBM	8.42	41.45	0.38	6.30
R_H_Kosk-Tiyatro	2	1	LGBM	29.23	42.60	−0.09	29.23
G_Tiyatro-H_Kosk	2	1	LGBM	10.19	16.81	0.11	9.39
R_Bursa-H_Kosk	2	2	RF	2.84	12.58	0.54	1.32
G_H_Kosk-Bursa	2	2	RF	2.20	8.16	0.67	1.23

Table A1. Cont.

Seg.-Dir.	Corr.	Seg. No.	Best Model	MAE (s)	RMSE (s)	R ²	ME (s)
R_Tokat–Bursa	2	3	LSTM	5.49	29.84	0.17	4.62
G_Bursa–Tokat	2	3	LGBM	2.51	15.44	0.06	2.14
R_Ulus–Tokat	2	4	LSTM	6.82	52.27	0.02	3.14
G_Tokat–Ulus	2	4	LSTM	5.90	29.59	0.16	4.29
R_Pekdemir–Ulus	3	1	LSTM	12.64	15.80	−0.12	−11.95
G_Ulus–Pekdemir	3	1	LGBM	7.05	8.90	0.25	−4.09
R_Kalp_Merkez–Pekdemir	3	2	RF	3.08	4.13	0.53	−0.48
G_Pekdemir–Kalp_Merkez	3	2	LR	6.01	7.99	0.46	2.82
R_Y.Tepe–Kalp_Merkez	3	3	LSTM	3.28	5.58	0.82	1.49
G_Kalp_Merkez–Y.Tepe	3	3	LSTM	3.28	4.39	0.63	0.09
R_Bagbasi–Yuksektepe	3	4	LSTM	2.67	4.50	0.61	1.04
G_Yuksektepe–Bagbasi	3	4	RF	4.02	5.26	0.42	0.24
R_Kent_Ormani–Bagbasi	3	5	LR	5.24	12.58	0.45	1.16
G_Bagbasi–Kent_Ormani	3	5	RF	3.26	4.11	0.53	−0.19
R_Kelleci–Kent_Ormani	3	6	RF	5.74	8.61	0.38	1.67
G_Kent_Ormani–Kelleci	3	6	LGBM	5.92	7.77	0.45	−0.98
R_Toki–Hal	4	1	RF	1.05	1.44	0.90	0.08
G_Hal–Toki	4	1	LSTM	23.85	31.43	−0.62	−23.84
R_1186Sk–Toki	4	2	LGBM	1.98	3.57	0.71	0.95
G_Toki–1186Sk	4	2	LGBM	3.03	6.37	0.60	1.58
R_1147Sk–1186Sk	4	3	RF	1.93	2.87	0.82	0.54
G_1186Sk–1147Sk	4	3	LGBM	2.73	6.32	0.58	−0.52
R_N_Kaynak–1147Sk	4	4	RF	7.72	12.67	−0.13	6.51
G_1147Sk–N_Kaynak	4	4	LSTM	0.68	0.89	0.87	0.14
R_Hilal_Cami–N_Kaynak	4	5	LR	15.47	17.94	0.42	1.18
G_N_Kaynak–Hilal_Cami	4	5	LGBM	9.54	12.29	0.38	6.05
R_Cakmak–Hilal_Cami	4	6	LGBM	1.64	2.85	0.73	−0.46
G_Hilal_Cami–Cakmak	4	6	RF	4.76	9.99	0.63	1.88
R_ASM–Cakmak	4	7	LSTM	1.84	2.67	0.86	0.92
R_Hallaclar–ASM	4	8	LR	23.38	27.31	−1.51	−23.29
R_Servergazi–Hallaclar	5	1	LSTM	2.84	5.08	0.65	−0.03
G_Hallaclar–Servergazi	5	1	RF	2.66	3.95	0.90	1.36
R_Aydem–Servergazi	5	2	LSTM	2.21	4.76	0.67	1.51
G_Servergazi–Aydem	5	2	LSTM	3.41	5.46	0.73	2.79
R_Adliye–Aydem	5	3	RF	26.14	28.44	−0.23	−25.41
R_Incilipinar–YSE	6	1	RF	3.32	5.35	0.89	1.30
G_YSE–Incilipinar	6	1	LSTM	21.77	26.16	0.22	−21.74
R_10Nisan–Incilipinar	6	2	LGBM	6.72	12.39	0.81	2.30
G_Incilipinar–10Nisan	6	2	LGBM	8.19	14.83	0.79	3.12
R_Sulukopru–10Nisan	6	3	LGBM	6.63	12.21	0.79	3.44
G_10Nisan–Sulukopru	6	3	LSTM	6.24	11.29	0.78	3.48
R_Cinar–Sulukopru	6	4	RF	3.92	5.77	0.94	−0.10
G_Sulukopru–Cinar	6	4	RF	4.92	15.48	0.57	0.25

References

1. Wu, Y.; Ye, Y.; Zeb, A.; Yu, J.J.; Wang, Z. Adaptive modeling of uncertainties for traffic forecasting. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 4427–4442. [\[CrossRef\]](#)
2. Yin, X.; Wu, G.; Wei, J.; Shen, Y.; Qi, H.; Yin, B. Deep learning on traffic prediction: Methods, analysis, and future directions. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 4927–4943. [\[CrossRef\]](#)
3. Yang, S.; Qian, S. Understanding and predicting travel time with spatio-temporal features of network traffic flow, weather and incidents. *IEEE Intell. Transp. Syst. Mag.* **2019**, *11*, 12–28. [\[CrossRef\]](#)
4. Tipi, R.; Şahin, H.; Doğru, Ş.; Zengin Bintaş, G.Ç. Comparison of the performances of regression-based machine learning algorithms in die design cost estimation. *Electron. Lett. Sci. Eng.* **2023**, *19*, 48–62.
5. Qiu, B.; Fan, W. Machine learning based short-term travel time prediction: Numerical results and comparative analyses. *Sustainability* **2021**, *13*, 7454. [\[CrossRef\]](#)
6. Shaygan, M.; Meese, C.; Li, W.; Zhao, X.G.; Nejad, M. Traffic prediction using artificial intelligence: Review of recent advances and emerging opportunities. *Transp. Res. Part C Emerg. Technol.* **2022**, *145*, 103921. [\[CrossRef\]](#)

7. Zheng, Z.; Ye, Y.; Zhu, Y.; Zhang, S.; Yu, J.J.Q. Data-driven methods for travel time estimation: A survey. In Proceedings of the IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), Bilbao, Spain, 24–28 September 2023. [\[CrossRef\]](#)
8. Lana, I.; Del Ser, J.; Velez, M.; Vlahogianni, E.I. Road traffic forecasting: Recent advances and new challenges. *IEEE Intell. Transp. Syst. Mag.* **2018**, *10*, 93–109. [\[CrossRef\]](#)
9. Abduljabbar, R.; Dia, H.; Liyanage, S. Machine learning models for traffic prediction on arterial roads using traffic features and weather information. *Appl. Sci.* **2024**, *14*, 11047. [\[CrossRef\]](#)
10. Kandiri, A.; Ghiasi, R.; Nogal, M.; Teixeira, R. Travel time prediction for an intelligent transportation system based on a data-driven feature selection method considering temporal correlation. *Transp. Eng.* **2024**, *18*, 100272. [\[CrossRef\]](#)
11. Pham, S.; Gerschberger, M.; Nistor, M.S.; Moll, M.; Cao, L.; Milani, R. Machine learning in vehicle travel time estimation: A brief technological perspective and review. In Proceedings of the 57th Hawaii International Conference on System Sciences, Waikiki, HI, USA, 3–6 January 2024; pp. 1409–1414. Available online: <https://hdl.handle.net/10125/106552> (accessed on 30 June 2026).
12. Ramezani, M.; Geroliminis, N. On the estimation of arterial route travel time distribution with Markov chains. *Transp. Res. Part B Methodol.* **2012**, *46*, 1576–1590. [\[CrossRef\]](#)
13. Patwary, A.U.Z.; Huang, W.; Lo, H.K. A link-to-link segment based metamodel for dynamic network loading. *Transp. Res. Part C Emerg. Technol.* **2021**, *130*, 103286. [\[CrossRef\]](#)
14. Zhang, Y.; Kong, X.; Zhou, W.; Liu, J.; Fu, Y.; Shen, G. A comprehensive survey on traffic missing data imputation. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 19252–19275. [\[CrossRef\]](#)
15. Xing, J.; Wu, W.; Cheng, Q.; Liu, R. Traffic state estimation of urban road networks by multi-source data fusion: Review and new insights. *Phys. A* **2022**, *595*, 127079. [\[CrossRef\]](#)
16. Varghese, V.; Chikaraishi, M.; Urata, J. Deep learning in transport studies: A meta-analysis on the prediction accuracy. *J. Big Data Anal. Transp.* **2020**, *2*, 199–220. [\[CrossRef\]](#)
17. Ayar, T.; Atlinar, F.; Güvensan, M.A.; Türkmen, H.İ. Long-term traffic flow estimation: A hybrid approach using location-based traffic characteristic. *Turk. J. Electr. Eng. Comput. Sci.* **2022**, *30*, 562–578. [\[CrossRef\]](#)
18. Koesdwiady, A.; Soua, R.; Karray, F. Improving traffic flow prediction with weather information in connected cars: A deep learning approach. *IEEE Trans. Veh. Technol.* **2016**, *65*, 9508–9517. [\[CrossRef\]](#)
19. Tsapakis, I.; Cheng, T.; Bolbol, A. Impact of weather conditions on macroscopic urban travel times. *J. Transp. Geogr.* **2013**, *28*, 204–211. [\[CrossRef\]](#)
20. Zechin, D.; Cybis, H.B.B. Probabilistic traffic breakdown forecasting through Bayesian approximation using variational LSTMs. *Transp. B Transp. Dyn.* **2023**, *11*, 1026–1044. [\[CrossRef\]](#)
21. Snehari, P.; Anvitha, B.C.S.S.; Kumar, A.M.; T, P.; Devi, M.M.Y.; Kumar, S. A comprehensive analysis on traffic prediction methods for real-world deployment: Challenges, cause and scope. In Proceedings of the International Conference on Communication, Computer Sciences and Engineering (IC3SE), Gautam Buddha Nagar, India, 9–11 May 2024; pp. 119–124. [\[CrossRef\]](#)
22. Kumar, B.A.; Singh, R.; Shaji, H.E.; Vanajakshi, L. Bus arrival time prediction: A comprehensive review. *IEEE Trans. Intell. Transp. Syst.* **2025**, *26*, 7362–7379. [\[CrossRef\]](#)
23. Petersen, N.C.; Rodrigues, F.; Pereira, F.C. Multi-output bus travel time prediction with convolutional LSTM neural network. *Expert Syst. Appl.* **2019**, *120*, 426–435. [\[CrossRef\]](#)
24. Liu, Y.; Lyu, C.; Zhang, Y.; Liu, Z.; Yu, W.; Qu, X. DeepTSP: Deep traffic state prediction model based on large-scale empirical data. *Commun. Transp. Res.* **2021**, *1*, 100012. [\[CrossRef\]](#)
25. Li, J.; Jiang, C.; Han, K.; Yu, Q.; Zhang, H. High-resolution spatiotemporal inference of urban road traffic emissions using taxi GPS and multi-source urban features data: A case study in Chengdu, China. *Urban Inform.* **2024**, *3*, 17. [\[CrossRef\]](#)
26. Abdellah, A.R.; Abdelmoaty, A.; Ateya, A.A.; Abd El-Latif, A.A.; Muthanna, A.; Koucheryavy, A. Accurate V2X traffic prediction with deep learning architectures. *Front. Artif. Intell.* **2025**, *8*, 1565287. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Shwartz-Ziv, R.; Armon, A. Tabular data: Deep learning is not all you need. *Inf. Fusion* **2022**, *81*, 84–90. [\[CrossRef\]](#)
28. Hewamalage, H.; Ackermann, K.; Bergmeir, C. Forecast evaluation for data scientists: Common pitfalls and best practices. *Data Min. Knowl. Discov.* **2023**, *37*, 788–832. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Kobayashi, K.; Alam, S.B. Explainable, interpretable, and trustworthy AI for an intelligent digital twin: A case study on remaining useful life. *Eng. Appl. Artif. Intell.* **2024**, *129*, 107620. [\[CrossRef\]](#)
30. Zahid, M.; Chen, Y.; Jamal, A.; Memon, M.Q. Short term traffic state prediction via hyperparameter optimization based classifiers. *Sensors* **2020**, *20*, 685. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Zahid, M.; Chen, Y.; Jamal, A.; Mamadou, C.Z. Freeway short-term travel speed prediction based on data collection time-horizons: A fast forest quantile regression approach. *Sustainability* **2020**, *12*, 646. [\[CrossRef\]](#)
32. Zheng, L.; Ma, H.; Wang, Z. Travel time estimation for urban arterials based on the multi-source data. *Sustainability* **2024**, *16*, 7845. [\[CrossRef\]](#)
33. Dong, H.; Pan, X.; Chen, X. Granularity optimization of travel trajectory based on Node2vec: A case study on urban travel time prediction. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 435. [\[CrossRef\]](#)

34. Li, Z.; Li, S.; Bamasag, O.O.; Alhothali, A.; Luo, X. Diversified regularization enhanced training for effective manipulator calibration. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *34*, 8778–8790. [CrossRef] [PubMed]
35. Chen, T.; Yang, W.; Li, S.; Luo, X. Data-driven calibration of industrial robots: A comprehensive survey. *IEEE/CAA J. Autom. Sin.* **2025**, *12*, 1544–1567. [CrossRef]
36. Karaoglan, M.E.; Murat, Y.S. Simulation-based optimization approach for effective corridor management: A case study of Denizli. *Akıllı Ulaşım Sist. Uygulamaları Derg.* **2025**, *8*, 237–255. [CrossRef]
37. TomTom International BV. Flow Segment Data. TomTom Developer Portal. 17 April 2025. Available online: <https://developer.tomtom.com/traffic-api/documentation/tomtom-maps/v1/traffic-flow/flow-segment-data> (accessed on 6 June 2026).
38. Visual Crossing Corporation. Timeline Weather API Documentation. Visual Crossing. 2026. Available online: <https://www.visualcrossing.com/resources/documentation/weather-api/timeline-weather-api/> (accessed on 6 June 2026).
39. Wang, L.; Chai, D.; Liu, X.; Chen, L.; Chen, K. Exploring the generalizability of spatio-temporal traffic prediction: Meta-modeling and an analytic framework. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 3870–3884. [CrossRef]
40. Chivers, B.D.; Wallbank, J.; Cole, S.J.; Sebek, O.; Stanley, S.; Fry, M.; Leontidis, G. Imputation of missing sub-hourly precipitation data in a large sensor network: A machine learning approach. *J. Hydrol.* **2020**, *588*, 125126. [CrossRef]
41. Transportation Research Board. *Highway Capacity Manual*, 6th ed.; Transportation Research Board: Washington, DC, USA, 2016; Chapter 17.
42. Chandra, S.; Kumar, U. Effect of lane width on capacity under mixed traffic conditions in India. *J. Transp. Eng.* **2003**, *129*, 155–160. [CrossRef]
43. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, NY, USA, 2009.
44. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
45. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
46. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3146–3154.
47. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
48. Géron, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed.; O'Reilly Media: Sebastopol, CA, USA, 2019.
49. Ma, X.; Tao, Z.; Wang, Y.; Yu, H.; Wang, Y. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transp. Res. Part C Emerg. Technol.* **2015**, *54*, 187–197. [CrossRef]
50. Zheng, F.; van Zuylen, H.; Liu, X. A methodological framework of travel time distribution estimation for urban signalized arterial roads. *Transp. Sci.* **2017**, *51*, 893–917. [CrossRef]
51. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.
52. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [CrossRef] [PubMed]
53. Lathuilière, S.; Mesejo, P.; Alameda-Pineda, X.; Horaud, R. A comprehensive analysis of deep regression. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2065–2081. [CrossRef] [PubMed]
54. Zhang, Y.; Haghani, A. A gradient boosting method to improve travel time prediction. *Transp. Res. Part C Emerg. Technol.* **2015**, *58*, 308–324. [CrossRef]
55. Ji, Y.; Geroliminis, N. On the spatial partitioning of urban transportation networks. *Transp. Res. Part B Methodol.* **2012**, *46*, 1639–1656. [CrossRef]
56. Liang, Z.; Wakahara, Y. Real-time urban traffic amount prediction models for dynamic route guidance systems. *EURASIP J. Wirel. Commun. Netw.* **2014**, *2014*, 85. [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.