Bi-Directional Isometric GRU for Urban Travel Time Prediction Using Deep Learning on Synthetic Global Positioning System Trajectories

Riki[1*] and Setyawan Widyarto[2]
[1]Information System Department, Faculty of Science and Technology, Buddhi Dharma University Banten, Indonesia 15115
[2]Computing Department, Faculty of Communication, Visual Art and Computing, Universiti Selangor Selangor, Malaysia 40000
Email: [1]riki@ubd.ac.id, [2]sety@unisel.edu.my

***Abstract*—Accurate short-term travel time prediction is crucial for optimizing urban transportation systems and enabling responsive Intelligent Transportation Systems (ITS). The research introduces a novel Deep Learning (DL) architecture, the Bi-Directional Isometric Gated Recurrent Unit (BDIGRU), to enhance predictive accuracy and resilience under congested traffic and variable conditions. BDIGRU uses isometric memory alignment to process temporal data in both forward and backward directions, enabling symmetric learning and balanced context integration. This is unlike traditional sequence models that are unidirectional. The researchers construct a high-resolution synthetic traffic dataset generated through stochastic simulation techniques to evaluate the model and replicate real-world urban traffic patterns in Jakarta, including variability in speed, volume, and congestion cycles. The dataset spans over 4 years, with more than 437,000 records captured at 5-min intervals. It includes multivariate traffic features such as average speed, vehicle volume, lane occupancy, and derived travel time. The performance of BDIGRU is benchmarked against five baseline models using four standard evaluation metrics. The results show that BDIGRU outperforms all the comparative models, achieving the lowest prediction errors (e.g., Mean Absolute Percentage Error (MAPE) = 7.2%, Symmetric Mean Absolute Percentage Error (SMAPE) = 8.0%). It also demonstrates superior robustness during high-variance traffic periods and congestion spikes. These findings underscore the architectural advantages of bidirectional and isometric temporal modeling for short-term traffic forecasting. Overall, BDIGRU offers a scalable and efficient framework suitable for real-time ITS deployment and highlights the value of synthetic datasets in developing and validating DL models for transportation analytics.**

***Index Terms*—Travel time prediction, Bi-Directional Isometric Gated Recurrent Unit (BDIGRU), Temporal Forecasting, Synthetic Traffic Data**

Received: April 20, 2025; received in revised form: Nov. 10, 2025; accepted: Nov. 11, 2025; available online: Aug. 26, 2026.
*Corresponding Author

I. INTRODUCTION

RAPID urbanization, rising vehicle ownership, and the proliferation of location-aware services have made urban mobility systems increasingly complex. These dynamics have contributed to pervasive congestion in urban and peri-urban regions worldwide, resulting in significant losses in travel efficiency, economic productivity, and environmental quality. The World Bank reports that traffic congestion in some Southeast Asian cities annually leads to losses of up to 2%–5% of GDP, underscoring the urgency of developing intelligent, adaptive, and predictive mobility solutions [1]. To address these systemic challenges, Intelligent Transportation System (ITS) has emerged as a cornerstone of urban development strategies worldwide. ITS integrates sensing technologies, communication networks, and Computational Intelligence (CI) to support real-time traffic management and improve transportation efficiency. Among the various functional components of ITS, travel time prediction stands out as one of the most critical, underpinning applications such as route guidance, fleet dispatching, traffic signal optimization, and Traveler Information Systems (TISs) [2].

Accurate travel time forecasting enables time-sensitive and cost-efficient decisions for stakeholders, including commuters, service providers, and logistics companies. However, achieving such accuracy is notoriously difficult, particularly in environments characterized by nonlinear dynamics. For instance, a 10-minute delay at a highway on-ramp due to an unexpected lane closure may ripple upstream, causing exponentially growing delays. Similarly, rain-induced congestion at 08:00 AM in a central business district may double the average travel time from 20 to 40 min, even when traffic flow appears nominally stable in preceding segments [3].

Traditional time series models, such as Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), and Exponential Smoothing, have long been employed for short-term travel time forecasting. Although these models are relatively simple to implement and interpret, they rely on strong assumptions of stationarity and linearity, which often do not hold in real-world traffic systems. Moreover, these approaches are limited in their ability to incorporate complex spatiotemporal interactions, particularly during congestion or incident-induced anomalies [3].

To overcome these limitations, many researchers have increasingly turned to Machine Learning (ML) and Deep Learning (DL) approaches, which can model complex nonlinear dependencies and learning patterns directly from large-scale traffic datasets. Recent advances in computing power and the availability of vehicle telemetry data make DL an attractive option for predicting travel time and traffic flow. Sequence models such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRUs) have shown remarkable performance in capturing temporal dependencies in sequential data [4].

ML methods have been explored to increase model flexibility and improve forecasting accuracy. Algorithms such as K-Nearest Neighbors (KNN), Support Vector Regression (SVR), and Random Forest have shown notable improvements in capturing nonlinear relationships and handling more diverse data distributions [5]. Nevertheless, these models still depend heavily on manual feature engineering, which can be time-consuming and susceptible to human bias. Moreover, they often struggle with learning long-range temporal dependencies, making them suboptimal for sequential prediction tasks where historical context is critical [6].

The introduction of DL methods, particularly RNN architectures, has provided a breakthrough in the modeling of sequential patterns in traffic data [7]. RNNs allow models to learn directly from the raw input timeline without requiring intensive preprocessing. However, standard RNNs encounter the vanishing gradient problem when processing long sequences, which limits their capacity to retain meaningful information over extended time lags. This limitation motivates the development of more sophisticated recurrent units such as LSTM and GRU, which aim to preserve memory and improve gradient flow across time steps [8]. To address this, LSTM and GRU models are introduced to maintain longer memory spans through gated control mechanisms [9]. Both LSTM and GRU have been widely adopted in traffic flow and travel time prediction due to their robustness in modeling nonlinear, time-varying dynamics [10].

Hybrid approaches that combine statistical models with Neural Networks (NNs), such as ARIMA-LSTM and SVR-GRU, have recently been introduced. These models aim to harness the interpretability of classical methods while benefiting from the learning capacity of NNs. Although hybrid models have demonstrated moderate improvements in accuracy, they tend to introduce additional architectural complexity. This makes them less scalable and difficult to deploy in real-time traffic environments, where computational efficiency is critical [11].

Despite their promise, conventional deep sequence models typically rely on unidirectional information flow, which processes input data from past to present but ignores future context. This approach can be suboptimal for travel time forecasting, where retrospective and anticipatory temporal patterns play essential roles. It may limit accuracy in contexts where future indicators, such as downstream congestion, influence current travel time [6]. This limitation has led to the adoption of Bidirectional RNNs (BiRNNs), which process sequences in both forward and backward temporal orders [10]. BiRNNs capture information in both forward and backward directions to enhance contextual learning [12]. However, despite their ability to use richer temporal context, conventional BiRNNs still have several weaknesses. They require higher computational cost than unidirectional RNNs because two recurrent passes are performed for each input sequence. In addition, standard BiRNNs remain vulnerable to unstable gradient propagation when processing long and highly fluctuating traffic sequences. More importantly, BiRNNs do not explicitly enforce temporal symmetry or equal contribution between forward and backward representations. This may cause imbalance in contextual learning, where one temporal direction dominates the prediction. Therefore, although BiRNNs improve contextual representation, their lack of isometric alignment can limit prediction stability in complex travel time forecasting tasks. However, in travel time scenarios, sequence inputs such as Eq. (1) are used. Ignoring such bidirectional dependencies can result in inaccurate or lagged predictions.

This limitation has led to the adoption of Bidirectional RNNs (BiRNNs), which process sequences in both forward and backward temporal orders [10]. BiRNNs capture information in both forward and backward directions to enhance contextual learning [12]. However, in travel time scenarios, sequence inputs such as Eq. (1):

$$X_t = \{speed_t, volume_t, occupancy_t\}_{(t=1)}^{T} \quad (1)$$

represents a sequence of traffic-condition data observed from time interval 1 to T. At each time step t, the data include the average vehicle speed ($speed_t$), traffic volume or the number of passing vehicles ($volume_t$), and detector occupancy, which indicates the percentage of time that the sensor is occupied by vehicles ($occupancy_t$).

Meanwhile, BiLTSM and BiGRU have demonstrated better performance in various transportation prediction tasks. Their ability to model bidirectional temporal dependencies allows them to generate richer contextual representations from sequential data [13]. However, most existing implementations of these bidirectional models do not enforce isometric alignment between forward and backward sequences. This lack of symmetry often results in uneven integration of temporal information and can degrade overall prediction quality [10].

The research aims to design and evaluate a DL architecture that captures bidirectional temporal dependencies while preserving strict temporal symmetry. It addresses the imbalance and distortion found in existing bidirectional models, such as BiGRU and BiLSTM, by introducing an isometric alignment mechanism that enforces equal-length sequence processing and balanced memory integration. This architectural refinement is the main novelty of the proposed Bi-Directional Isometric Gated Recurrent Unit (BDIGRU) model and distinguishes it from conventional recurrent approaches.

The BDIGRU extends the standard GRU through bidirectional temporal processing, capturing both forward ($t \rightarrow T$) and backward ($T \rightarrow t$) dependencies in time-series data [14]. Its isometric alignment mechanism applies equal-length input sequences and consistent time intervals in both temporal directions. This symmetry maintains balanced memory representations across past and future contexts, reduces bias in feature extraction, and improves prediction stability. BDIGRU also improves the processing of temporal concurrency and causal feedback, making it suitable for traffic scenarios involving cyclical congestion, signalized intersections, and time-of-day effects. The architecture uses separate memory cells for ascending and descending sequences, followed by a fully connected fusion layer that combines directional features into a unified output vector. This structure supports robust prediction across traffic conditions with different volatility levels.

In addition to sequential models, graph-based DL methods, such as Graph Convolutional Networks (GCNs) and Attention-Based Spatial Graph Neural Networks (AST-GCN), have also gained popularity [15]. These models emphasize spatial dependencies across sensor networks and road segments, making them suitable for large-scale deployments [16]. However, graph-based models often require dense sensor coverage and complete traffic topology information, which may not always be available or possible to obtain in real-world conditions. This makes sequential DL models, such as BDIGRU, more practical for applications that rely on limited or streaming data sources, such as mobile Global Positioning System (GPS) logs or low-resolution IoT inputs [17].

The key advantage of BDIGRU lies in its ability to integrate past and future temporal information simultaneously [12, 14]. By bidirectionally processing time series data, the model can better detect patterns that other unidirectional models often miss [10]. Consequently, it becomes more robust in handling noise, outliers, and abrupt changes in traffic conditions [18].

In parallel, another line of research addresses the challenges of data scarcity and limited traffic prediction reproducibility. Previous research has employed synthetic or simulated datasets to mitigate these issues, which allow for controlled experimentation and benchmarking of model performance [19]. While synthetic data offers the advantage of reproducibility and scenario flexibility, its utility depends heavily on how accurately it replicates real-world traffic dynamics and statistical distributions [11].

As a result, the research also addresses a major empirical limitation in ITS research: the scarcity of labeled real-time traffic datasets. Many existing models rely on proprietary or localized datasets that are not publicly available, limiting reproducibility, generalizability, and methodological rigor. To reduce this constraint, a synthetic traffic dataset is developed using stochastic simulation, periodic seasonality, and congestion modeling to emulate realistic urban traffic conditions.

The synthetic dataset is constructed to represent multiple traffic states, including free-flow, saturated flow, and stop-and-go waves, across multiple road segments and time intervals. For example, a sequence in Eq. (2):

$$X_t = [avg\_speed_t^{(1)}, avg\_speed_t^{(2)}, \ldots, avg\_speed_t^{(N)}] \quad (2)$$

represents the spatial traffic-speed feature vector at time step t. Each element $avg\_speed_t^{(1)}$ denotes the average vehicle speed observed at the i-th road segment, where $i = 1, 2, \ldots, N$, and N is the total number of road segments considered in the network. It has N = 10 road segments, updated every 5 min, serving as input to the model to predict travel time $\varphi_t + \Delta$, where Δ = 30 minutes. Using this synthetic dataset, the researchers evaluate the predictive performance of BDIGRU against several benchmark models, including ARIMA, LSTM, GRU, Deep-RNN, and Bidirectional LSTM. The researchers adopt four standard error metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Symmetric Mean Absolute Percentage Error (SMAPE) to quantify accuracy under different congestion regimes.

The main research problem is the extent to which the proposed BDIGRU model can simultaneously capture bidirectional temporal dependencies within equal-length windows to improve the accuracy of short-term travel time prediction under varying traffic conditions. The research examines the BDIGRU architecture by addressing three critical subproblems. The first subproblem is whether bidirectional isometric processing can improve memory discrimination compared with traditional unidirectional GRU or LSTM models. Second, the performance of BDIGRU under simulated congestion shocks relative to baseline models is evaluated. Third, the extent to which synthetic traffic data can replicate the predictive complexity found in real-world traffic scenarios is investigated. Equation (3) provides a mathematical formulation of the proposed model:

$$X = \{x_1, \ldots, x_2\}, x_t \in \mathbb{R}^d, \tag{3}$$

where $x_t \in \mathbb{R}^d$ denotes an input sequence containing T observations. Each vector x_t is a d-dimensional feature vector at time step t that d represents the number of traffic-related features included in the model.

Overall, the literature reflects a consistent progression from statistical approaches to DL-based models, with increasing emphasis on mechanisms such as bidirectional memory encoding, attention-based architectures, and the use of synthetic validation datasets [19]. Nevertheless, an unresolved methodological gap persists. Specifically, the integration of bidirectional sequence learning with structural isometry has received limited attention.

II. Research Method

This section presents the mathematical formulation and architectural design of the BDIGRU model, which is developed to improve travel time prediction under dynamic and uncertain traffic conditions [14]. Although the research primarily focuses on the proposed BDIGRU architecture, the overall methodology includes a complete experimental pipeline. Synthetic traffic data generation, selection and normalization of relevant multivariate features, and comparative evaluation using multiple baseline models are performed. The procedure involves structured data preprocessing, without overlap, temporal splitting for training, validation, and testing, and the use of consistent training configurations across all models. Each component, from dataset construction to performance evaluation, is designed to ensure transparency, fair comparison, and reproducibility. Figure 1 illustrates the workflow of this research, which explains every research step from the problem-definition stage to evaluation and conclusion.

The key components are included to ensure stability and adaptability in real-world applications. BDIGRU builds upon the conventional GRU framework by introducing an isometric alignment mechanism, denoted as $\int$. This component is designed to enhance memory retention across time steps, improve temporal feature learning, and provide greater robustness when dealing with nonstationary patterns in traffic data [20]. Through this structural augmentation, BDIGRU aims to preserve temporal symmetry and reduce distortion in sequential information processing.

A. Problem Definition

Let $X = x_1, x_2, \ldots, x_T$, where $x_t \in \mathbb{R}^d$ denotes the multivariate input sequence at time ttt, such as average speed, traffic volume, and occupancy from sensors on multiple road segments [2]. The goal is to predict the short-term travel time at a future step $T + \Delta$, denoted by $y_T + \Delta \in \mathbb{R}^d$, using both historical and prospective temporal context [7]. The objective is to learn a function $\int : \mathbb{R}^T \times d \to \mathbb{R}$ such that it maps a multivariate input sequence x_t into a predicted travel time y_T. As demonstrated in Eq. (4), this mapping is expressed as:

$$y_T + \Delta = \int (X), \tag{4}$$

where $y_T + \Delta$ denotes the estimated travel time at time step t, and x_t represents a feature vector from multiple road segments consisting of average speed, volume, and occupancy [4].

The GRU is an RNN variant specifically designed to address the challenges of learning long-term dependencies in sequential data. Unlike traditional RNNs that suffer from vanishing gradients when managing long sequences, the GRU introduces gating mechanisms, namely, the update and reset gates, which regulate the flow of information, allowing the model to retain or discard temporal features across time steps dynamically. This mechanism enables more effective learning of both short- and long-term temporal patterns while

Cite this article as: Riki and S. Widyarto, "Bi-directional isometric GRU for urban travel time prediction using deep learning on synthetic global positioning system trajectories", CommIT Journal 20(2), 347–361, 2026.

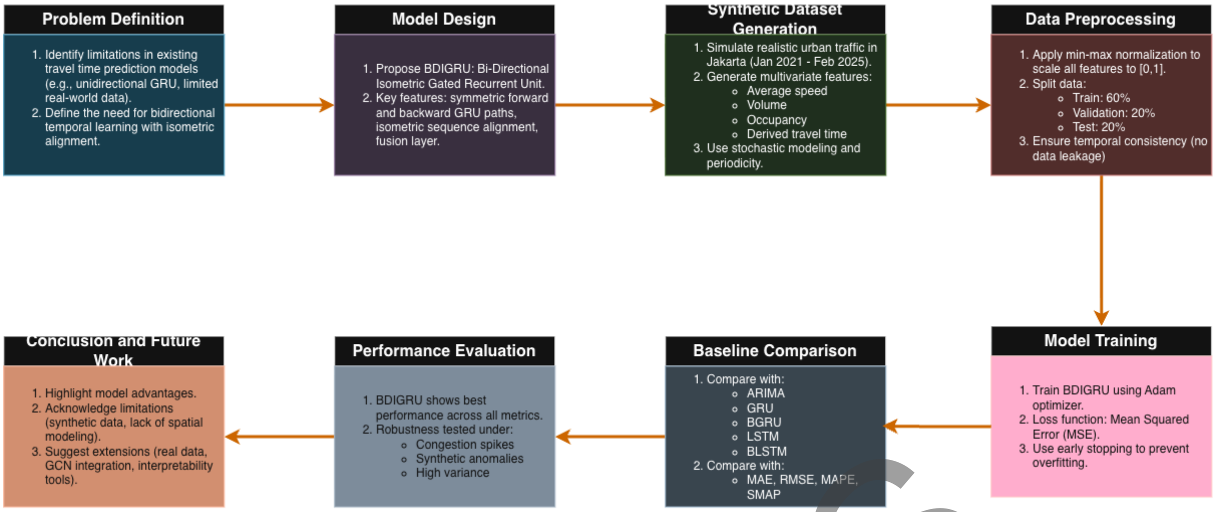

Fig. 1. Research workflow for Bi-Directional Isometric Gated Recurrent Unit (BDIGRU)-based prediction of urban travel time. Note: Autoregressive Integrated Moving Average (ARIMA), Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Symmetric MAPE (SMAPE).

keeping the model architecture simple compared to its predecessor, the LSTM network [19].

At each time step t, the GRU computes two primary gates. The reset gate, denoted as r_t, determines the extent to which the previous hidden state $h_(t-1)$ should be ignored when processing the new input x_t. This relationship is captured in Eq. (5) [18]:

$$r_t = \sigma(W_r[h_{t-1}, x_t] + b_r) \tag{5}$$

where W_r and U_t are weight matrices and b_r is the bias term. Then, $\sigma(.)$ is the sigmoid activation function [7]. Similarly, the update gate, denoted as ztz_tzt, controls the balance between the previous memory and the newly generated state, as expressed in Eq. (6):

$$u_t = \sigma(W_u[h_{t-1}, x_t] + b_u). \tag{6}$$

In this context, W_u, u_t, and b_u follow analogous roles as in the reset gate, facilitating the selective retention of temporal information. Then, the candidate hidden state h_t representing the potential update to the memory is calculated using the reset gate as shown in Eq. (7):

$$h_t = \tanh(W_c[r_t \bigodot h_{t-1}, x_t] + b_c). \tag{7}$$

In this formulation, $\bigodot$ denotes the Hadamard (element-wise) product [7]. The reset gate r_t modulates how much of the previous hidden state h_{t-1} contributes to the candidate state. The weights W_c, h_t, b_c and bias are learnable parameters. The tanh function introduces non-linearity to the state transition. Finally, the actual hidden state h_t at time step t is computed by interpolating between the previous hidden state and the candidate state, weighted by the update gate. This is expressed in Eq. (8):

$$h_t = (1 - u_t) \bigodot h_{t-1}, u_t \bigodot h_t). \tag{8}$$

These equations define how past memory is updated with current inputs at each timestep [21].

GRU adaptively balances the retention of past information and the incorporation of new inputs through its gating mechanism, allowing the memory state to respond to gradual and abrupt changes in traffic patterns with lower computational complexity than LSTM. Building on this mechanism, the BDIGRU processes each input sequence in two symmetric directions, forward and backward, using independent parameters [14]. The forward hidden state $h_t^{\rightarrow}$ captures information from earlier time steps, whereas the backward hidden state $h_t^{\leftarrow}$ captures contextual information from later positions within the historical input window. By integrating these complementary representations, BDIGRU improves its ability to learn complex temporal traffic dynamics while maintaining a lightweight architecture suitable for short-term travel time prediction. The forward hidden state in BDIGRU is computed by passing the input at time t and the hidden state from the next time step $t + 1$ (due to backward recurrence in bidirectional computation), as shown in Eq. (9):

$$h_t^{\rightarrow} = \text{GRU} \rightarrow (x_t, h_{t+1}^{\rightarrow}). \tag{9}$$

Similarly, the hidden state in the backward direction

Cite this article as: Riki and S. Widyarto, "Bi-directional isometric GRU for urban travel time prediction using deep learning on synthetic global positioning system trajectories", CommIT Journal 20(2), 347–361, 2026.

TABLE I
SAMPLE SYNTHETIC TRAFFIC DATASET (JAKARTA, 2021–2025).

Timestamp	Avg. Speed (km/h)	Volume	Occupancy (%)	Travel Time (min)
01/01/2021 00:00	47.50	335	45.80	3.07
01/01/2021 08:15	25.30	650	91.20	7.94
03/07/2022 17:40	58.20	410	35.50	2.59
14/11/2023 06:30	18.70	702	83.00	8.41
28/02/2025 22:45	39.10	510	50.70	4.21

is computed as in Eq. (10):

$$h_t^{\leftarrow} = \text{GRU} \leftarrow (x_t, h_{t+1}^{\leftarrow}). \tag{10}$$

These computations reflect the dual temporal exploration in BDIGRU, where future and past contexts are symmetrically coded using dated recurrent dynamics. This bidirectional processing captures contextual dependencies that are more vibrant and richer, which is important for predicting travel time accurately in volatile urban traffic conditions. Next, the concatenated vector combining both hidden states is expressed in Eq. (11):

$$h_t^{\rightarrow} = [h_t^{\rightarrow}; h_t^{\leftarrow}]. \tag{11}$$

This concatenated vector h_t encodes bidirectional contextual information, which is particularly beneficial in transportation scenarios where current travel time is affected not only by historical congestion but also by anticipated downstream traffic states. Such fusion allows the model to build a comprehensive view of the traffic system at each time step.

The last step in the BDIGRU architecture involves projecting the combined hidden representation to the predicted output through a fully connected output layer, as formulated in Eq. (12):

$$y_T + \Delta = \sigma(W_o.h_t + b_o) \tag{12}$$

Here, $y_T + \Delta$ represents the predicted travel time at forecast horizon $T + \Delta$, W_o, and b_o are the weight matrix and bias of the output layer, and σ denotes an activation function. Depending on the regression or classification task, σ can be a linear identity function or a nonlinear activation (e.g., Rectified Linear Unit (ReLU) or $\tanh$). This structure ensures that BDIGRU not only captures bidirectional temporal dynamics but also performs high-capacity mapping from hidden representations to forecasted travel time [13].

B. Dataset and Experimental Setup (Synthetic Dataset Design)

A high-resolution synthetic traffic dataset was generated to emulate traffic telemetry in the Jakarta metropolitan area from January 2021 to February 2025 at five-minute intervals. The dataset includes average speed, traffic volume, lane occupancy, and travel time to represent dynamic urban traffic conditions. Average speed is generated between 5 and 80 km/h, traffic volume ranges from 50 to 1,000 vehicles per interval, and lane occupancy varies from 5% to 95% following periodic congestion patterns. Travel time, measured in minutes, is calculated using a nonlinear congestion-sensitive function, as defined in Eq. (13):

$$\begin{aligned} travel\_time_t = 10 \times (\frac{1.2}{speed_t}) + \\ 0.8.\frac{occupancy_t}{100} + \epsilon_t, \end{aligned} \tag{13}$$

where $\epsilon_t \sim N(0,\ 0.05)$ represents random measurement noise. This setup helps the dataset to better reflect real traffic patterns under different conditions [2].

The travel time function reflects the inverse relationship between average speed and travel time, while incorporating lane occupancy as a proxy for congestion intensity, where higher occupancy generally produces longer travel durations. To improve realism, random variability is added to emulate traffic disturbances, including signal delays, unexpected events, and fluctuations in vehicle arrivals and departures [4]. Table I presents a sample of the synthetic dataset, illustrating the input features, their measurement scales, and the variation of traffic conditions across time steps.

Before training, all traffic features are rescaled to the range [0,1] using min-max normalization to reduce differences in feature magnitude, improve numerical stability during gradient-based optimization, and support faster model convergence [14]. The normalized data are then organized into multivariate time series sequences that enable efficient batching and integration into the deep learning training pipeline. As illustrated in Fig. 2, BDIGRU consists of an input layer, a BDIGRU layer, a concatenation layer, and a fully connected output layer. The input layer receives traffic features such as average speed, volume, and occupancy. At the same time, the forward and backward GRU paths learn complementary temporal representations from earlier and later positions within the historical input window. These representations are concatenated and mapped by the output layer to produce the predicted travel time. This design enables BDIGRU to capture complex and

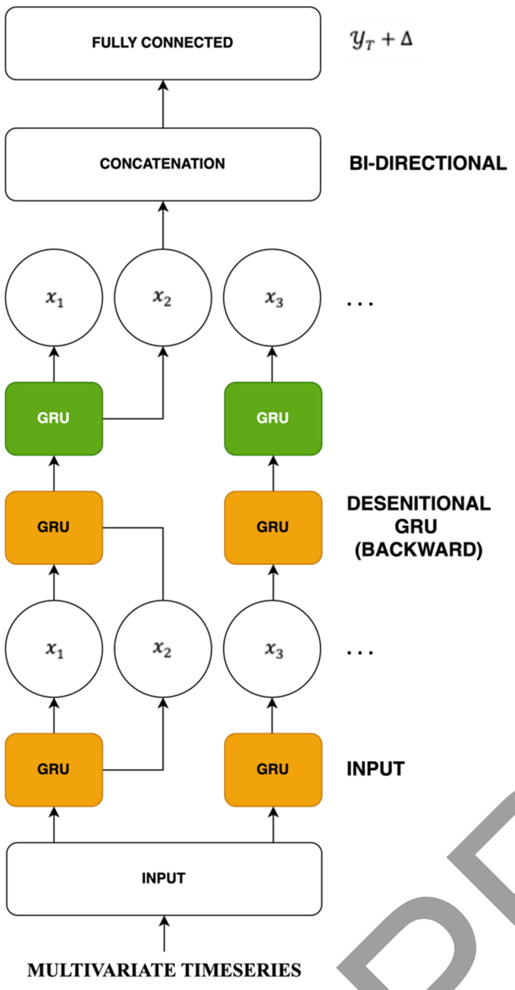

Fig. 2. The Bi-Directional Isometric Gated Recurrent Unit (BDI-GRU) architecture.

nonstationary traffic patterns while maintaining an efficient architecture for short-term travel time prediction.

This configuration allows the BDIGRU model to effectively learn and represent long-range temporal dynamics while maintaining structural symmetry. The isometric alignment reduces information imbalance between temporal paths and improves the model's predictive accuracy under fluctuating traffic conditions [13, 21]. The training of the BDIGRU model involves optimizing a set of parameters $\Theta = \{W_r, W_u, W_c, W_o, b_r, b_u, b_c, b_o \ldots\}$, which include the weight matrices and bias terms associated with the forward and backward GRU layers, as well as the output layer. The objective is to minimize the prediction error between the true travel time y_i and the model's estimated output $\widehat{y}_i$. This is formalized using the MSE loss function, which is defined as shown in Eq. (14):

$$L(\theta) = \frac{1}{N} \sum_{i-1}^{N} (y_i - \widehat{y}_i)^2, \tag{14}$$

where N represents the total number of training samples. The MSE is chosen due to its sensitivity to large errors and its differentiability, making it well-suited for gradient-based optimization methods [11]. The goal is to find the optimal parameter set $\Theta*$ that minimizes this loss over the training dataset [14].

BDIGRU is trained using backpropagation through time with an optimizer such as Stochastic Gradient Descent (SGD) or Adam. Model parameters are randomly initialized, and the training data $\{(X_i, y_i)\}_{i=1}^{N}$ are shuffled and processed in mini batches during each epoch. For every input sequence, the forward GRU processes observations from $t = 1$ to T, whereas the backward GRU processes the same historical window from T to 1. Their hidden states, $h_t^{\rightarrow}; h_t^{\leftarrow}$, are concatenated into a unified representation $h_t^{\rightarrow} = [h_t^{\rightarrow}; h_t^{\leftarrow}]$, which is passed to a fully connected layer to predict travel time $y_T + \Delta$. The loss is computed by comparing the predicted and actual travel times, and the model parameters are updated iteratively to minimize this error. Early stopping based on validation loss is applied to reduce overfitting and select the best-performing model [22].

C. Data in the Research

The complete synthetic Jakarta traffic dataset comprises 437,473 timestamped records spanning the period from January 2021 to February 2025. It is provided in comma-separated values (.csv) format and has been made publicly accessible for academic and experimental use (https://doi.org/10.5281/zenodo.21304849). This dataset is specifically designed to support research in ITS. It has multivariate traffic features such as speed, volume, and occupancy, and simulates realistic urban traffic dynamics. These characteristics make the dataset suitable for benchmarking predictive models, confirming algorithmic robustness, and simulating traffic management scenarios using data-driven methods [15].

D. Data Splitting

The complete dataset is partitioned into three subsets: a training set comprising 60% of the data (262,483 records), a validation set comprising 20% (87,495 records), and a testing set comprising 20% (87,495 records). This partitioning is executed using a non-overlapping rolling window strategy to maintain temporal integrity across splits and prevent information leakage between the training and evaluation phases. Such a strategy ensures that the model is assessed on future data, and it has not been exposed to during training, thereby enhancing the robustness of performance evaluation [17, 23].

E. Baseline Models for Comparison

BDIGRU is compared with five baseline models: ARIMA, GRU, BiGRU, LSTM, and BiLSTM. All neural models use identical training settings, including the Adam optimizer with a learning rate of 0.001, 64 hidden units, and a dropout rate of 0.2. The models are implemented in PyTorch 2.0 and trained on an NVIDIA RTX GPU, while grid search and early stopping are used to select suitable hyperparameters and prevent overfitting [24]. Meanwhile, performance is evaluated using MAE, RMSE, MAPE, and SMAPE, measuring absolute error, squared error, relative percentage error, and symmetric percentage error, respectively [24]. These metrics offer a comprehensive evaluation by capturing both absolute and relative prediction errors. Such dual assessments allow more robust accuracy measurement of the model in various traffic conditions, including high volatility and traffic jams [11, 22].

MAE calculates the average of the absolute differences between predicted and actual values [25]. It provides a straightforward measure of how far predictions deviate from actual observations, regardless of direction. A lower MAE indicates better predictive accuracy. However, MAE treats all errors equally and does not emphasize larger errors more than smaller ones. The formula is shown in Eq. (15):

$$\mathrm{MAE} = \frac{1}{N}\sum_{N}^{i-1}|y_i - \widehat{y}_i|, \tag{15}$$

where y_i is the actual travel time for the i-th observation, $\widehat{y}_i$ is the corresponding predicted travel time, and N is the total number of observations.

RMSE computes the square root of the average of the squared differences between predicted and actual values. Because it squares the errors before averaging, RMSE penalizes larger errors more heavily. This situation makes it particularly useful when large deviations are more critical. A lower RMSE reflects higher prediction precision, especially in the presence of outliers [25]. The formula is shown in Eq. (16):

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{N}^{i-1}(y_i - \widehat{y}_i)^2}. \tag{16}$$

MAPE expresses the prediction error as a percentage of the actual value. It offers a relative measure of accuracy, making it easier to interpret across datasets with varying scales [25]. However, it can become unstable or undefined when actual values are remarkably close to zero, as the formula involves division by the actual value (see Eq. (17)):

$$\mathrm{MAPE} = \frac{100\%}{N}\sum_{N}^{i-1}|\frac{y_i - \widehat{y}_i}{y_i}|. \tag{17}$$

SMAPE is a modified version of MAPE that uses the average of the predicted and actual values in the denominator [25]. This adjustment symmetrizes the error and reduces the impact of extreme values or small denominators. It provides a more balanced percentage error, minimizing bias between overestimation and underestimation. It is shown in Eq. (18):

$$\mathrm{SMAPE} = \frac{100\%}{N}\sum_{N}^{i-1}\frac{|y_i - \widehat{y}_i|}{|y_i| + |\widehat{y}_i|/2} \tag{18}$$

III. Results and Discussion

The model input includes a sliding window of $T-12$ time steps (i.e., 1-hour history at 5-minute intervals), each holding 4 dimensions: average speed, volume, occupancy, and lagged travel time [7]. These features are selected based on their empirical relevance in finding spatiotemporal congestion levels, and their codependency supports modeling nonlinear interactions in deep learning settings. To ensure consistent feature magnitudes, feature scaling is applied using min-max normalization as expressed in Eq. (19):

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}, \tag{19}$$

where x_{min} and x_{max} represent the minimum and maximum values of that feature in the dataset, respectively, and x' is the normalized value used as model input.

The dataset includes four input features recorded at each time step, namely average speed (km/h), traffic volume (vehicles per five-minute interval), lane occupancy (%), and lagged travel time. These features are chosen for their fundamental relevance in traffic theory, along with the empirical significance demonstrated in prior predictive modeling studies. Together, they capture both instantaneous and temporal traffic dynamics, enabling the model to learn complex patterns associated with travel time fluctuations [14].

A. Performance Results of the Bi-Directional Isometric Gated Recurrent Unit (BDIGRU) Model

This section presents the numerical results of BDIGRU in a clear and organized way. The presentation includes a combination of descriptive statistics, error metric evaluations, and visual illustrations to assess

Cite this article as: Riki and S. Widyarto, "Bi-directional isometric GRU for urban travel time prediction using deep learning on synthetic global positioning system trajectories", CommIT Journal 20(2), 347–361, 2026.

TABLE II
MODEL EVALUATION RESULTS (TESTING SET).

Model	MAE	RMSE	MAPE (%)	SMAPE (%)
Autoregressive Integrated Moving Average (ARIMA)	1.25	1.80	12.50	13.80
Gated Recurrent Unit (GRU)	0.92	1.40	9.30	10.40
Bidirectional GRU (BiGRU)	0.89	1.35	9.00	10.00
Long Short-Term Memory (LSTM)	0.91	1.38	9.10	10.10
Bidirectional LSTM (BiLSTM)	0.87	1.32	8.70	9.60
Bi-Directional Isometric GRU (BDIGRU)	0.74	1.10	7.20	8.00

Note: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Symmetric MAPE (SMAPE).

TABLE III
COMPARISON OF BI-DIRECTIONAL ISOMETRIC GATED RECURRENT UNIT (BDIGRU) WITH PREVIOUS MODELS IN TRAVEL TIME PREDICTION.

Study/Model	Architecture	Dataset Type	MAPE (%)	Key Limitation/Note	Advantage Compared to Prior Work
This research (BDIGRU)	BDIGRU	Synthetic (Jakarta, 2021–2025)	7.2	None observed; robust under high variance and congestion	It outperforms GRU, BiGRU, LSTM, and BiLSTM with bidirectional isometry
6	BiLSTM	Large-scale traffic logs	9.6	No isometric alignment between sequences	BDIGRU achieves $>$ 2.4% lower MAPE
10	CNN-GRU-LSTM (Hybrid)	Real traffic + sensors	8.9	High computational complexity and training time	BDIGRU is simpler and faster and has equal or better accuracy
5	SVR-GRU (Hybrid)	Simulated + empirical	10.2	Requiring manual feature engineering; limited to short-term prediction	BDIGRU has end-to-end learning and better generalization
15	AST-GCN (Graph Model)	Real-time GCN input	$\sim$8.5	Requiring dense spatial topology (sensors, road networks)	BDIGRU is more scalable in sparse data environments
Traditional ARIMA (Benchmark)	Statistical Time Series	Univariate (Baseline)	12.5	Assuming stationarity; poor under nonlinear conditions	BDIGRU handles nonlinear, bidirectional time dependencies

Note: Autoregressive Integrated Moving Average (ARIMA), Gated Recurrent Unit (GRU), Bidirectional GRU (BiGRU), Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Mean Absolute Percentage Error (MAPE), Convolutional Neural Network (CNN), Support Vector Regression (SVR), and Attention-Based Spatial Graph Neural Networks (AST-GCN).

model performance across multiple conditions. Five tables and five figures are presented to support the analysis and to make it easier to compare the model with the baseline models.

The performance of the BDIGRU model is benchmarked against several baseline models: ARIMA, GRU, BiGRU, LSTM, and BiLSTM. Table II presents the comparative results using four metrics: MAE, RMSE, MAPE, and SMAPE. The BDIGRU model consistently achieves the lowest error rates, showing its superior predictive capacity under dynamic traffic patterns.

As shown in Table II, BDIGRU outperforms all baseline models across the four evaluation metrics, achieving an MAE of 0.74, RMSE of 1.10, MAPE of 7.2%, and SMAPE of 8.0%. The low MAE indicates that BDIGRU produces the smallest average difference between predicted and actual travel times. In contrast, the low RMSE shows that it also reduces larger prediction errors that may occur during sudden congestion or highly variable traffic conditions [24]. Compared with BiLSTM, the strongest baseline model, BDIGRU reduces RMSE from 1.32 to 1.10, representing an improvement of approximately 16.7%. In terms of relative error, its MAPE is 1.8% lower than BiGRU and 1.5% lower than BiLSTM, demonstrating more accurate predictions across different traffic intensities. The lowest SMAPE value further indicates a balanced reduction of underestimation and overestimation errors. Overall, these results suggest that the bidirectional isometric design enables BDIGRU to learn temporal traffic patterns more effectively and deliver stable short-term travel time predictions [11, 14, 15, 22].

To provide contextual grounding for the observed performance, Table III presents a comparative summary of the BDIGRU model against several existing travel time prediction models. This comparison highlights differences in architectural approaches, evaluation metrics, and model robustness across varying traffic scenarios. By synthesizing empirical outcomes from prior studies, it serves as a benchmark reference to underscore the relative advantages of the proposed BDIGRU framework in terms of prediction accuracy, generalizability, and computational feasibility.

The experimental results confirm that BDIGRU outperforms ARIMA, GRU, BiGRU, LSTM, and BiLSTM

TABLE IV
DESCRIPTIVE STATISTICS OF TRAFFIC FEATURES.

Variable	Mean	Std.	Min	Q1	Median	Q3	Max
Average Speed (km/h)	39.80	14.57	5.00	29.58	39.36	49.98	79.99
Volume	298.50	100.20	50.00	225.10	293.40	366.90	999.70
Occupancy (%)	49.95	19.90	5.02	35.26	49.82	64.55	94.99
Travel Time (min)	4.96	1.75	1.43	3.69	4.80	6.03	10.78

across all evaluation metrics. The model achieves a MAPE of 7.2% and a SMAPE of 8.0%, indicating higher predictive accuracy than conventional recurrent architectures. These results suggest that bidirectional learning with isometric alignment improves temporal representation and supports more reliable travel time forecasting under changing traffic conditions.

Previous research [6] has reported a MAPE of approximately 9.6% for BiLSTM, compared with 10.1% for LSTM and 10.4% for GRU. BDIGRU reduces the MAPE by 2.4 percentage points relative to BiLSTM. This improvement indicates that symmetrical memory alignment helps the model learn historical and anticipatory traffic patterns more effectively than conventional bidirectional networks without structural isometry.

Hybrid models, such as ARIMA-LSTM and SVR-GRU, have also improved forecasting accuracy [5, 9]. However, these approaches often increase computational complexity and reduce scalability for real-time deployment. BDIGRU retains a relatively simple sequence-based architecture while remaining robust during congestion spikes, traffic noise, and high-variance conditions [6, 11].

Graph-based models such as AST-GCN perform well when dense sensor networks and complete road topology are available [15]. Their application becomes difficult in environments with sparse spatial data. Because BDIGRU is topology-independent, it can operate without extensive sensor coverage and is suitable for ITS environments with limited data sources [16]. Overall, BDIGRU offers accurate, scalable, and robust travel time prediction. It addresses asymmetrical sequence processing by balancing forward and backward information within equal temporal windows during model training.

B. Statistics of Traffic Features

The average vehicle speed in the dataset is 39.80 km/h, with a standard deviation of 14.57 km/h. This variation demonstrates that there is a traffic segment with smooth flow along with another very congested segment. A minimum recorded speed of 5.00 km/h likely corresponds to severe traffic jams or signal-induced delays. In contrast, the maximum speed of approximately 80 km/h reflects conditions on urban expressways during low congestion periods. This wide range of speed values enhances the dataset's ability to simulate diverse real-world scenarios. According to descriptive statistics in Table IV, the interquartile range for speed falls between 29.21 km/h (Q1) and 50.40 km/h (Q3). This indicates that 50% of traffic conditions are concentrated in this moderate speed range. This finding is consistent with inter-city mobility patterns under mixed traffic conditions [2].

Regarding traffic volume, the dataset reports a mean of 298.5 vehicles per 5-minute interval, with a substantial standard deviation of 100.2 vehicles. It indicates fluctuating demand across different time windows. The peak volume is almost 1,000 vehicles, reflecting critical weight, such as morning and evening peak hours. Such variability is essential for testing model robustness under high-load and transition conditions [26].

Occupancy is defined as the proportion of roadway space occupied by vehicles, averaging 49.95%. The minimum value of 5.02% captures free-flow conditions, while the maximum value of 94.99% reflects near-saturation, typical of severe congestion. This spread across low- and high-density states provides a nuanced representation of temporal load changes on urban infrastructure [11].

The target variable, which is the simulated travel time, has a mean of 4.96 minutes with a standard deviation of 1.75 minutes. Its distribution is notably right-skewed, with the third quartile (Q3) at 6.03 minutes, indicating that most trips fall within a moderate time range. However, outliers reaching up to 10.78 minutes highlight extreme congestion events, likely caused by incidents or bottlenecks [4]. Overall, these statistical indicators confirm the dataset's ability to represent complex traffic dynamics. It includes both predictable patterns and extreme conditions, which is crucial for evaluating time-series prediction models under realistic and stress-testing scenarios [15].

C. Traffic Dynamics in Average Speed and Travel Time

Figure 3 shows marked daily speed fluctuations, with recurring declines during simulated morning and evening peaks as traffic volume and road occupancy increase. This pattern supports the realism of the synthetic dataset in representing congestion-related speed

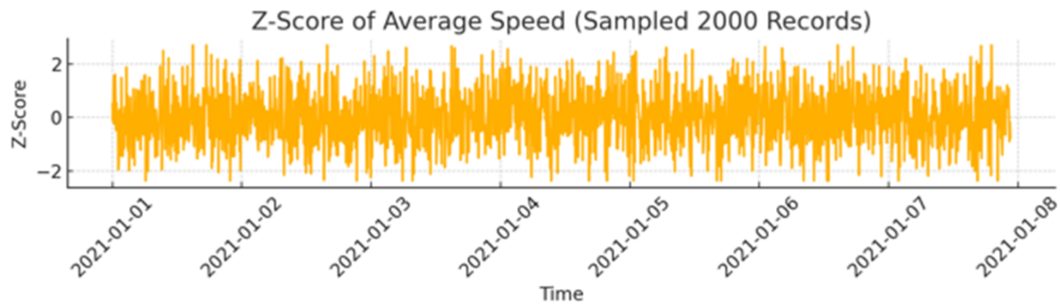

Fig. 3. Z-score of average speed.

reductions. The standardized Z-score representation further highlights how speed values deviate from their average level over time, allowing abnormal speed drops and temporary recovery periods to be identified more clearly. Negative Z-score values indicate periods when vehicle speeds fall below the mean, which commonly reflects congested or unstable traffic conditions. In contrast, positive values represent smoother traffic flow with speeds above the average. These variations demonstrate that the synthetic data capture both regular daily mobility cycles and short-term disruptions, making it suitable for evaluating BDIGRU's ability to learn temporal patterns under fluctuating urban traffic conditions.

Next, Z-score analysis identifies extreme travel-time deviations. Table V lists the five largest outliers recorded from 2021 to 2025. These events represent severe congestion or disruptions and provide important cases for testing model robustness under high-variance traffic conditions. All five outliers have Z-scores above 3.8, placing them within the top 0.01% of travel-time values observed under a Gaussian assumption [27]. These extreme values represent severe congestion or traffic disruptions caused by simulated incidents, such as lane closures, peak-hour saturation, or sudden traffic breakdowns [28, 29]. The largest outlier occurred on February 1, 2023, at 14:50, with a Z-score of 4.22. The recorded travel time was 4.22 standard deviations above the mean, indicating an unusually severe congestion event [2]. Several anomalies also appear outside normal rush hours, including at 04:05 and 23:45. This pattern suggests that disruptions may arise from random events or structural irregularities [11]. These findings support the need for robust forecasting models. BDIGRU applies symmetric bidirectional learning to capture forward and backward temporal patterns, allowing it to maintain stable predictions under extreme traffic conditions [14].

Figure 4 presents the normalized Z-score distribution of travel time from 2,000 synthetic records collected

TABLE V
TOP FIVE Z-SCORE OUTLIERS OF TRAVEL TIME.

Timestamp	Z-Score Travel Time
01/02/2023 14:50	4.22
23/11/2023 23:45	3.99
08/01/2024 04:05	3.91
11/08/2024 23:45	3.90
01/03/2021 08:05	3.89

between January 1 and 8, 2021. Most values remain within ± 2, indicating relatively stable traffic conditions. However, several sharp spikes reveal statistical outliers associated with extreme congestion or unexpected disruptions. These anomalies occur intermittently and outside peak periods, suggesting that traffic disturbances are not always linked to daily cycles. This pattern supports the need for robust forecasting models. BDIGRU captures both historical and anticipatory temporal dependencies through bidirectional processing. This structure allows the model to maintain prediction accuracy under volatile and anomalous traffic conditions.

D. *Traffic Dynamics in Volume and Occupancy*

Figures 5 and 6 depict the z-score visualizations for traffic volume and occupancy, respectively. These figures illustrate how both variables fluctuate over time within the simulated traffic environment. Temporal peaks are clearly observed in both plots. Notably, these peaks align with typical urban rush hour periods, such as morning (07:00–09:00) and evening (16:00–19:00) intervals. This alignment strongly indicates that the synthetic dataset successfully emulates realistic urban traffic dynamics. The z-score transformation allows for standardized comparisons across features with different scales. In this context, it helps to highlight abnormal increases or decreases in traffic flow and road usage intensity. These visual patterns further confirm the quality and behavioral accuracy of the generated dataset.

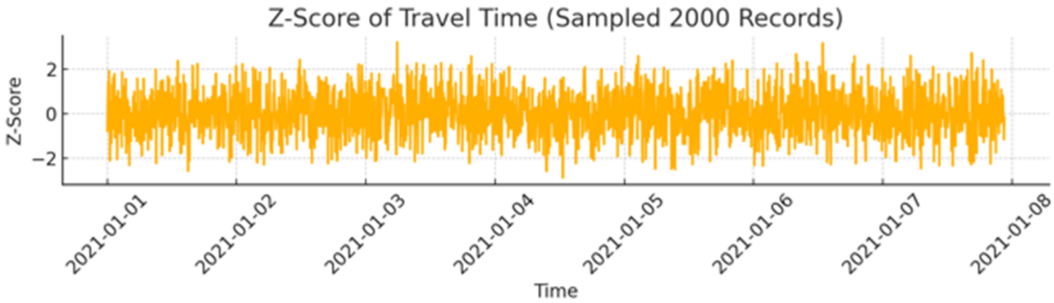

Fig. 4. Z-score of travel time.

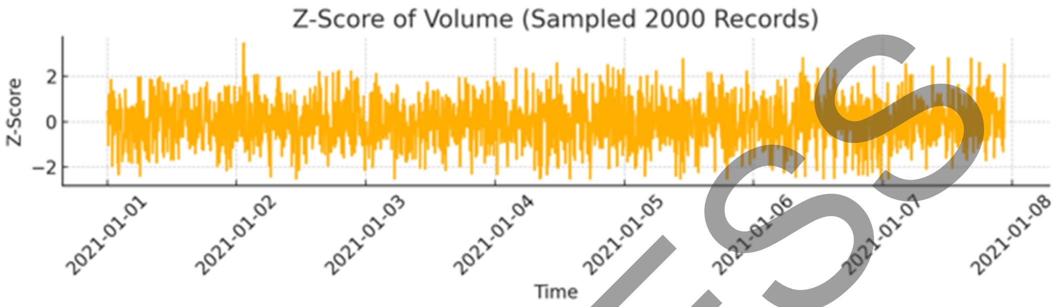

Fig. 5. Z-score of traffic volume.

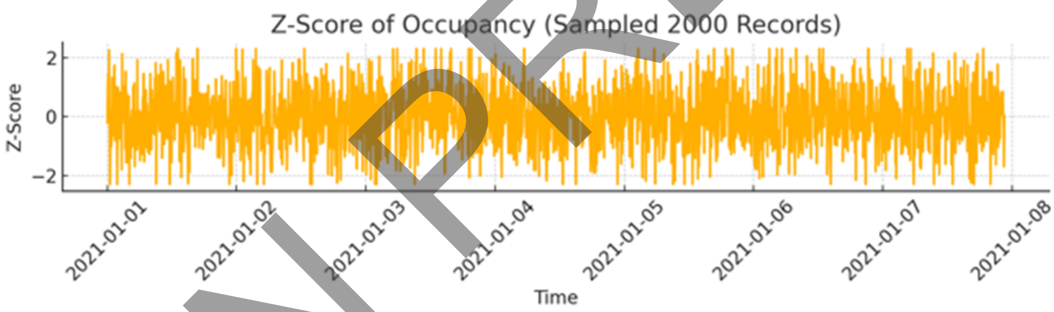

Fig. 6. Z-score of occupancy.

They demonstrate that the synthetic traffic conditions are not randomly assigned but are instead constructed to mimic observable real-world phenomena under both normal and congested states.

Figure 5 shows the Z-score distribution of traffic volume for 2,000 samples collected from January 1–8, 2021. Most values fall within ± 2, indicating generally stable traffic flow, while sharp positive and negative spikes reflect abrupt changes in vehicle counts. These fluctuations represent simulated peak-hour surges and random disturbances. The heavier positive tail indicates brief periods of intense traffic accumulation. BDIGRU addresses this variability through bidirectional sequence processing, learning from preceding and subsequent traffic states to improve prediction accuracy under nonstationary conditions.

Figure 6 presents the Z-score distribution of lane occupancy from 2,000 records collected between January 1 and 8, 2021. Values range from approximately -2.5 to +2.5, reflecting frequent shifts between free-flow and congested conditions. Recurring spikes indicate high temporal volatility caused by signal changes, lane merging, and sudden demand increases. This pattern confirms that occupancy is a sensitive congestion indicator. BDIGRU captures these fluctuations through bidirectional and isometric temporal processing, enabling more accurate estimation of occupancy-related travel-time changes than unidirectional models.

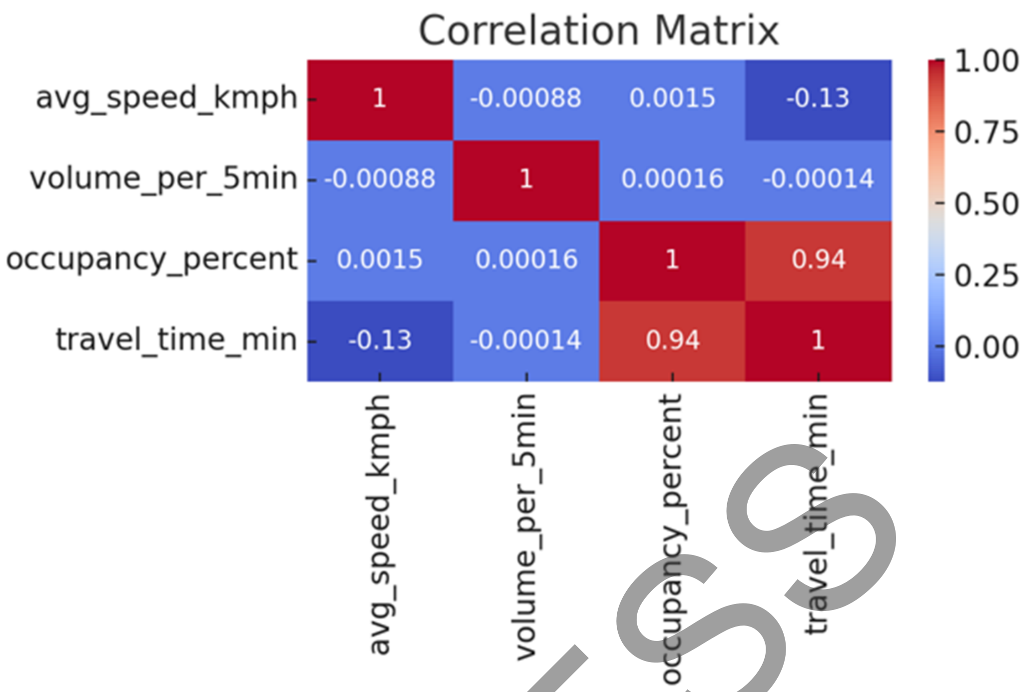

Fig. 7. Results of correlation matrix.

E. Result of Correlation Matrix

Figure 7 presents the Pearson correlation heatmap for key traffic variables. Speed and travel times have a strong negative correlation of approximately -0.78, indicating that lower vehicle speeds correspond to longer travel times during congestion or disruptions. In contrast, occupancy and travel times show a strong positive correlation of about 0.82. Higher lane occupancy reflects denser traffic and longer travel duration. These relationships confirm that the synthetic dataset preserves realistic traffic behavior. They also support the inclusion of speed, volume, and occupancy as BDIGRU input features. The model achieves high one-step-ahead travel-time accuracy, particularly under severe congestion, sudden speed changes, irregular volume spikes, and non-recurrent incidents.

IV. Conclusion

The research proposes the BDIGRU, a DL architecture for robust short-term travel-time prediction under complex urban traffic conditions. By leveraging bidirectional sequence modeling and isometric memory alignment, BDIGRU captures temporal dependencies within the historical input window and achieves superior forecasting performance across MAE, RMSE, MAPE, and SMAPE. Experiments on a large-scale synthetic dataset representing Jakarta's metropolitan traffic show that BDIGRU consistently outperforms ARIMA, GRU, BiGRU, LSTM, and BiLSTM. The model is particularly effective during highly volatile and congested traffic periods, achieving a MAPE of 7.2% and an SMAPE of 8.0%. The results indicate improved prediction accuracy and stability in dynamic urban mobility scenarios.

Nevertheless, several limitations remain. The synthetic dataset, although designed to reflect realistic traffic behavior, does not include actual sensor-derived data, signal phases, road geometry, or recorded traffic events, which limits the generalizability of the findings. In addition, BDIGRU focuses on temporal dependencies and does not explicitly model spatial interactions among connected road segments. It is also not evaluated in real-time or online-learning environments. Future work should validate the model using empirical data from loop detectors, GPS trajectories, or mobile sensors, and incorporate graph-based approaches such as GCNs or Graph Attention Networks (GANs) to capture spatiotemporal dependencies. Interpretability techniques, robustness tests using noisy or incomplete data, and ablation studies are also needed to clarify the contribution of each architectural component and support trustworthy deployment in adaptive ITS applications.

BDIGRU should be validated using real-world traffic data from diverse road types, cities, weather conditions, and traffic-demand levels to assess its generalizability under sensor noise, missing data, incidents, road works, and sudden congestion. The lightweight model may also be integrated into edge-based ITS platforms for low-latency traffic signal control, congestion detection, dynamic route guidance, and logistics planning, while considering trade-offs among prediction accuracy, inference time, memory use, energy consumption, and communication cost. Interpretability methods such as Shapley additive explanations (SHAP) values, temporal attention, and feature-importance analysis can be applied to identify the contributions of speed, volume, occupancy, and historical time steps to each prediction. Comprehensive ablation studies should compare BDIGRU with unidirectional GRU, conventional Bi-GRU, and variants without the isometric component. At the same time, robustness tests should evaluate performance under noisy, incomplete, delayed, and imbalanced traffic data.

Acknowledgement

The authors received no financial support for the research.

Author Contribution

The data that support the findings of this research are openly available in Zenodo at https://doi.org/10.5281/zenodo.21304849.

Data Availability

Developed the research concept, formulated the BDIGRU framework, and designed the experimental procedure, R.; Conducted the main comparative analysis and interpreted the model performance, R.; Prepared the original manuscript draft, discussion, and conclusion, R.; Performed validation, prepared visualizations, reviewed the manuscript, and completed the final editing, R.; Generated, organized, and curated the synthetic Jakarta traffic dataset, S. W.; Prepared the preprocessing pipeline, analytical scripts, model implementation, and computational resources, S. W.; Conducted model training, hyperparameter tuning, performance evaluation, and robustness testing, S. W.; and Contributed to the methodology, experimental setup, results, tables, figures, and manuscript revision, S. W.

References

[1] The World Bank, “Planning, connecting & financing cities–Now,” 2013. [Online]. Available: https://bit.ly/4cBQH4r

[2] Y. An, M. Li, X. Lin, F. He, and H. Yang, “Space-time routing in dedicated automated vehicle zones,” *Transportation Research Part C: Emerging Technologies*, vol. 120, 2020.

[3] D. Rezaei, I. Aghayan, and F. Hadadi, “Studying perturbations and wave propagations by lane closures on traffic characteristics based on a dynamic approach,” *Physica A: Statistical Mechanics and its Applications*, vol. 566, 2021.

[4] R. Deshmukh, D. Sun, K. Kim, and I. Hwang, “Temporal logic learning-based anomaly detection in metroplex terminal airspace operations,” *Transportation Research Part C: Emerging Technologies*, vol. 126, 2021.

[5] D. Liu, Z. Wang, Y. Liu, and F. E. Alsaadi, “Extended Kalman filtering subject to random transmission delays: Dealing with packet disorders,” *Information Fusion*, vol. 60, pp. 80–86, 2020.

[6] H. Rao, S. Xu, X. Hu, J. Cheng, and B. Hu, “Augmented skeleton based contrastive action learning with momentum LSTM for unsupervised action recognition,” *Information Sciences*, vol. 569, pp. 90–109, 2021.

[7] A. Sarraf and S. Khalili, “An upper bound on the variance of scalar multilayer perceptrons for log-concave distributions,” *Neurocomputing*, vol. 488, pp. 540–546, 2022.

[8] J. B. Florindo, Y. S. Lee, K. Jun, G. Jeon, and M. K. Albertini, “VisGraphNet: A complex network interpretation of convolutional neural features,” *Information Sciences*, vol. 543, pp. 296–308, 2021.

[9] V. Singh, S. K. Sahana, and V. Bhattacharjee, “A novel CNN-GRU-LSTM based deep learning model for accurate traffic prediction,” *Discover Computing*, vol. 28, pp. 1–25, 2025.

[10] J. Tang *et al.*, “Deep spatio-temporal dependent convolutional LSTM network for traffic flow prediction,” *Scientific Reports*, vol. 15, pp. 1–15, 2025.

[11] W. Guo, Q. Zeng, H. Duan, W. Ni, and C. Liu, “Process-extraction-based text similarity measure for emergency response plans,” *Expert Systems with Applications*, vol. 183, 2021.

[12] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” *IEEE transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[13] L. L. Zhang, D. U. Gang, W. U. Jun, and M. A. Yujie, “Joint production planning, pricing and retailer selection with emission control based on Stackelberg game and nested genetic algorithm,”

Expert Systems with Applications, vol. 161, 2020.

[14] C. Y. T. Chen, E. W. Sun, M. F. Chang, and Y. B. Lin, "Enhancing travel time prediction with deep learning on chronological and retrospective time order information of big traffic data," *Annals of Operations Research*, vol. 343, no. 3, pp. 1095–1128, 2024.

[15] W. Xiao, H. Liu, Z. Ma, and W. Chen, "Attention-based deep neural network for driver behavior recognition," *Future Generation Computer Systems*, vol. 132, pp. 152–161, 2022.

[16] M. Seppecher, L. Leclercq, A. Furno, D. Lejri, and T. V. Da Rocha, "Estimation of urban zonal speed dynamics from user-activity-dependent positioning data and regional paths," *Transportation Research Part C: Emerging Technologies*, vol. 129, 2021.

[17] H. Han, J. Teng, J. Xia, Y. Wang, Z. Guo, and D. Li, "Predict high-frequency trading marker via manifold learning," *Knowledge-Based Systems*, vol. 213, 2021.

[18] Y. Kim, H. y. Tak, S. Kim, and H. Yeo, "A hybrid approach of traffic simulation and machine learning techniques for enhancing real-time traffic prediction," *Transportation research part C: emerging technologies*, vol. 160, 2024.

[19] M. S. Ansari, "A single-layer asymmetric RNN with low hardware complexity for solving linear equations," *Neurocomputing*, vol. 485, pp. 74–88, 2022.

[20] F. Orozco, H. Wen, P. P. De Gusm, J. Wahlström, and M. Luo, "Federated learning for traffic flow prediction with synthetic data augmentation," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 7, pp. 445–459, 2026.

[21] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, "Dive into deep learning." [Online]. Available: https://d2l.ai/index.html

[22] V. Bonato and C.-S. Bouganis, "Class-specific early exit design methodology for convolutional neural networks," *Applied Soft Computing*, vol. 107, 2021.

[23] W. Shu, K. Cai, and N. N. Xiong, "Research on strong agile response task scheduling optimization enhancement with optimal resource usage in green cloud computing," *Future Generation Computer Systems*, vol. 124, pp. 12–20, 2021.

[24] V. Cerqueira, L. Torgo, and I. Mozetič, "Evaluating time series forecasting models: An empirical study on performance estimation methods," *Machine Learning*, vol. 109, pp. 1997–2028, 2020.

[25] P. Manandhar, H. Rafiq, E. Rodriguez-Ubinas, and T. Palpanas, "New forecasting metrics evaluated in prophet, random forest, and long short-term memory models for load forecasting," *Energies*, vol. 17, no. 23, pp. 1–30, 2024.

[26] G. Li, Y. Chen, Y. Wang, P. Nie, Z. Yu, and Z. He, "City-scale synthetic individual-level vehicle trip data," *Scientific Data*, vol. 10, pp. 1–18, 2023.

[27] T. Thomas, W. Weijermars, and E. Van Berkum, "Variations in urban traffic volumes," *European Journal of Transport and Infrastructure Research*, vol. 8, no. 3, pp. 251–263, 2008.

[28] Y. Wang *et al.*, "Exploiting topic-based adversarial neural network for cross-domain keyphrase extraction," in *2018 IEEE International Conference on Data Mining (ICDM)*. Singapore: IEEE, Nov. 17–20, 2018, pp. 597–606.

[29] J. B. Sheu, "Microscopic traffic behaviour modelling and simulation for lane-blocking arterial incidents," *Transportmetrica A: Transport Science*, vol. 9, no. 4, pp. 335–357, 2013.