Bus-Passenger-Flow Prediction Model Based on WPD, Attention Mechanism, and Bi-LSTM

Yulong Pei, Songmin Ran *, Wanjiao Wang and Chuntong Dong

School of Civil Engineering and Transportation, Northeast Forestry University, Harbin 150040, China
* Correspondence: ransm@nefu.edu.cn

Abstract: The prediction of the bus passenger flow is crucial for efficient resource allocation, frequency setting, and route optimization in bus transit systems. However, it remains challenging for a single model to simultaneously capture the time-series data of the bus passenger flow with periodicity, correlation, and nonlinearity. Aiming at the complex volatility possessed by the time-series data of the bus passenger flow, a new hybrid-strategy bus-passenger-flow prediction model based on wavelet packet decomposition, an attention mechanism, and bidirectional long–short-term memory is proposed to improve the accuracy of bus-passenger-flow prediction. The differences between this study and the existing studies are as follows: Firstly, this model combines decomposition strategies and deep learning. Wavelet packet decomposition can decompose the original data into a series of smoother data components, allowing the model to be more adequate in capturing the temporal characteristics of passenger-flow data. And the model can consider the information after the predicted moment via backward computation. In addition, the model is equipped with the ability to focus on important features by incorporating an attention mechanism to minimize the interference of irrelevant information. Bus-passenger-flow prediction experiments are conducted using the Harbin bus-passenger-flow dataset as an example. The experimental results show that the model proposed in this paper can obtain more accurate bus-passenger-flow prediction results than the five baseline models can obtain.

Keywords: urban traffic; public transit; deep learning; decomposition; hybrid model

1. Introduction

Public transportation is a crucial constituent of the transportation system, which can alleviate urban traffic congestion and mitigate CO_2 emissions. Improving the quality of public transportation services can elevate the usage rate of public transportation and gradually make ecofriendly public transportation the main mode of urban travel. However, it is arduous to provide the general public with high-quality public transportation for many cities in developing nations [1]. For example, unreliable bus services, such as those involving irregular departure schedules or bus crowding [2], can discourage the public from choosing bus travel. In addition, with the proliferation of on-demand transportation services like Uber and Lyft, bus traffic has been flatlining in recent years [3]. Bus services' effective operation has been adopted to enhance the quality of service in the bus system, which is an effective means to increase bus ridership.

The accurate and proactive prediction of the bus passenger flow facilitates transit agencies to implement dynamic control strategies, making them more effective and responsive to passengers' travel needs, which improves the service level of daily bus operations and attracts more urban travelers to choose bus transit as a mode of travel. In addition, the timely and reliable prediction of the passenger flow can help transit agencies optimize their bus schedules and reduce their operating costs [4–6].

Essentially, the prediction of the bus passenger flow is a time-series problem. Statistics-based prediction methods are the classical methods often applied to time-series prediction,

but these methods are difficult to utilize when dealing with the complex nonlinear relationships that bus-passenger-flow time series possess [7–9]. Machine learning and deep learning methods have a wide range of application prospects in traffic data prediction, and these methods can effectively deal with the nonlinear relationships of the data with a high prediction accuracy [10,11]. The bidirectional long–short-term memory model in the deep learning approach is able to capture both the forward and backward time-series data features of the traffic-flow data and is able to provide good prediction results [12].

The above methods have made some positive progress in the predicting performance, but there is still room for improvement in the accuracy of the bus-passenger-flow prediction methods. There are numerous variables that affect the bus passenger flow, such as the time of day, the traffic conditions, and the weather, which make the bus passenger flow nonstationary and unpredictable [13]. In the case of time-series bus-passenger-flow data containing a mix of linear and nonlinear information, neither traditional statistical models nor deep learning models can accomplish satisfactory prediction results by themselves. Thus, considering the complexity of the problem of predicting time-series data, many researchers have chosen to incorporate a decomposition strategy when making predictions. They argue that the extraction of the complex features of traffic time-series data is facilitated by decomposing the raw data using data decomposition methods [14,15].

It is worth considering whether the performance of the bus-passenger-flow prediction model can be further enhanced by introducing a decomposition strategy in order to obtain more accurate and realistic prediction results. Therefore, this paper proposes a bus-passenger-flow prediction model based on wavelet packet decomposition (WPD), an attention mechanism (ATT), and bidirectional long–short-term memory (Bi-LSTM). WPD can decompose the original transit-passenger-flow time-series data into a series of smoother components, which helps the model to better extract the features of time-series data. In particular, the model also combines an attention mechanism with Bi-LSTM. The Bi-LSTM model is able to capture the correlation of time-series data in both the forward and backward directions, while an attention mechanism can help the Bi-LSTM model focus on important feature information. Based on the above ideas, this paper constructs the WPD-ATT-BiLSTM model to capture the changing patterns of the time-series data of the bus passenger flow and to make predictions.

The main contributions of this paper are as follows:

- A new hybrid-strategy bus-passenger-flow prediction model, the WPD-ATT-BiLSTM model, is proposed in this paper. Processing bus-passenger-flow time-series data with obvious fluctuation characteristics through WPD can make the data loaded into the Bi-LSTM model smoother, which helps the model to better perform data feature capturing in both directions. The inclusion of an attention mechanism allows the model to focus on the impact of the important features of the data on the prediction results.

- To address the fluctuating characteristics of the bus-passenger-flow data, WPD is used to decompose them into smoother components. For the nonlinearity and periodicity of the bus-passenger-flow data, combining the Bi-LSTM model with an attention mechanism enhances the ability of the model to capture the changing pattern of the passenger flow.

- The effectiveness of the proposed model is validated using the Harbin transit-passenger-flow dataset. The results show that the accuracy of the results of the WPD-ATT-BiLSTM model is higher than that of the results of either of the benchmark models in both single-step and multistep prediction. The use of decomposition strategies and attention mechanisms significantly improves the prediction performance of the model.

This study is organized as follows: Section 2 provides a review of the existing time-series forecasting methods in the field of transport. Section 3 details the framework of the model proposed in this paper and the principles of the mathematical model used. Section 4 describes the dataset used for model validation, conducts prediction experiments, and provides detailed experimental results. Section 5 analyzes the experimental results and

explains the reasons for the excellent performance of the model proposed in this paper. Section 6 concludes the paper.

2. Literature Review

With the rapid adoption of sensors and the internet in practice, abundant spatial-temporal traffic data can be recorded [16]. There has been a great amount of interest in learning how to mine the space–time rules of big data to improve the prediction of traffic time-series data. The existing common methods for traffic data prediction can be classified into two categories, namely, statistic methods and machine learning methods. Statistics-based prediction methods are the classical methods commonly used for passenger-flow prediction, which mainly include regression analysis, the moving average, and the Kalman filter. Zheng et al. (2020) [17] presented a model based on sparse regression for predicting the traffic flow. Cai et al. (2019) [18] found that real traffic data contain non-Gaussian noises and thus used a Kalman filter model along with the maximum correlation entropy to predict the traffic flow. Moving average models are useful for time-series analysis, and they identify patterns and predict future trends by separating the long-term changes and seasonal cycles in historical data [19]. The autoregressive integrated moving average (ARIMA) model is the basis for most moving average models [7]. Shahriari et al. (2020) [20] presented bringing together bootstrapping with the conventional parametric ARIMA model to create an ensemble of ARIMA models for predicting the traffic flow.

Machine learning prediction methods are widely used due to the nonlinear characteristics of multiple types of traffic data, such as traffic-flow data and passenger-flow data, because they are better able to fit such complicated data [21,22]. Chen et al. (2020) [23] introduced artificial neural networks (ANNs) to predict the traffic flow over different time spans. Wang et al. (2019) [24] proposed a regression framework for short-term traffic-flow prediction that utilizes support vector regression (SVR) and can perform automatic parameter tuning. Liu et al. (2019) [25] developed a random forest (RF) model to predict the passenger flow and evaluated the impact of different input feature combinations on the prediction accuracy. Sun et al. (2021) [26] and Lu et al. (2023) [27] used extreme gradient boosting (XGBoost) to predict the traffic volume on the highway.

In recent years, the emergence and widespread adoption of deep learning models using machine learning have caused a significant stir in the transportation industry [28,29]. In addition, a growing corpus of research employs models of deep learning to predict the passenger flow. For example, Lin et al. (2019) [30] proposed an end-to-end deep-learning-based model to improve the accuracy and stability of air-traffic-flow prediction. Chen et al. (2021) [31] designed a framework for predicting the traffic flow in urban road networks using deep learning. Nagaraj et al. (2022) [32] developed a method based on deep learning with long–short-term memory (LSTM), recurrent neural networks (RNNs), and greedy hierarchical algorithms for predicting the bus passenger flow. Du et al. (2020) [33] designed an LSTM network model for predicting how people move through urban traffic. Han et al. (2019) [34] created a hybrid optimization of the LSTM model to predict the flow of bus passengers. The current corpus of research indicates that the LSTM model, within the domain of deep learning, is particularly adept at processing time-series data exhibiting long-term correlation and, consequently, has been widely adopted for traffic time-series prediction [35].

Bi-LSTM was proposed based on a single LSTM model and has been applied to traffic time-series data prediction. Abduljabbar et al. (2021) [36] used the Bi-LSTM model for short-term traffic prediction on three different highways. Zhai et al. (2022) [37] utilized the Bi-LSTM model to predict the short-term traffic flow on urban roads, and the prediction experimental results showed that the prediction accuracy of the Bi-LSTM model is higher than that of the single LSTM model. The Bi-LSTM model extracts the forward and backward features of time-series data at the same time, which causes the model to have better prediction results than the traditional LSTM model when dealing with strongly periodic time-series data [38].

To address the limitations of a single model and to capitalize on the strengths of various models, some researchers are integrating single models into hybrid models to enhance the prediction accuracy [39]. Glisovic et al. (2016) [40] demonstrated a composite passenger prediction model utilizing the genetic algorithm (GA) and artificial neural networks (ANNs). Xu et al. (2017) [41] presented a road-traffic-flow prediction method based on the ARIMA and the Kalman filter. Lin et al. (2021) [42] designed a short-term traffic-flow prediction model combining the ARIMA model and the generalized autoregressive conditional heteroskedasticity (GARCH) model. Li et al. (2023) [43] combined empirical mode decomposition (EMD), the sample entropy (SE), and the kernel extreme learning machine (KELM) to predict the short-term bus passenger flow. In addition, some studies have considered the integration of an attention mechanism and the LSTM model to predict traffic time-series data. The experimental results show that adding an attention mechanism module to the LSTM model can effectively improve the prediction performance of the model [44]. Incorporating an attention mechanism into the prediction model can help the model better capture the nonlinear features of traffic time-series data. Attention mechanisms work by calculating the importance of the input features of traffic time-series data at different moments. The weight values of the input features are determined according to the importance of the input features so that the model focuses on important feature information and reduces the interference of irrelevant information [45].

Common decomposition methods for traffic data include the wavelet transform (WT), empirical mode decomposition, seasonal adjustment, variational mode decomposition, and intrinsic time-scale decomposition. Currently, the WT has received significant attention [46]. The WT decomposes the signal into its high-frequency and low-frequency components, transforming it from nonstationary to stationary and causing the subsequent analysis to be less complicated. Khandelwal et al. (2015) [47] have demonstrated that the WT can improve the accuracy of time-series forecasts. Diao et al. (2019) [48] used the discrete wavelet transform (DWT) to decompose the traffic volume sequence into an allocation component and a number of detailed components and then predicted them using a tracking model and a Gaussian process model, respectively. Zhu et al. (2021) [49] developed a short-term traffic-flow prediction method based on the WT and a multidimensional Taylor network (MTN), where the WT was used to decompose the traffic flow to improve the prediction accuracy.

The WT method can decompose a signal profile into a series of profiles of different frequencies. WPD was developed based on the WT, and it is a more refined decomposition method compared to the WT [50]. Bus-passenger-flow time-series data can be considered to be fluctuating nonlinear signals. The use of WPD enables the information of the different frequency bands contained in passenger-flow time-series data to be effectively extracted and enables a detailed analysis to be performed based on the information characteristics of the different frequency bands.

3. Methods

3.1. The Entire Process of the Proposed Model

Passenger-flow prediction is essentially a time-series prediction problem. To make the prediction model better at extracting feature information from the time-series data, WPD was used to decompose the original transit-passenger-flow data. WPD is able to take raw bus-passenger-flow data, which is volatile and complex, and decompose it into a series of smoother components. Specifically, after the WPD decomposition process, the raw bus-passenger-flow data was decomposed into two types of components: an approximate component at low frequency and a detailed component at high frequency. Prediction models are better able to extract features and capture patterns of time-series data when dealing with smooth components. In addition, the proposed model combined an attention mechanism with the Bi-LSTM model. The Bi-LSTM model extracts the features of bus-passenger-flow data components in both the forward and backward directions simultaneously. And adding the attention mechanism can make the Bi-LSTM model pay

more attention to important time-series data features in expectation of achieving better prediction accuracy. The framework of the model constructed in this paper is shown in Figure 1.
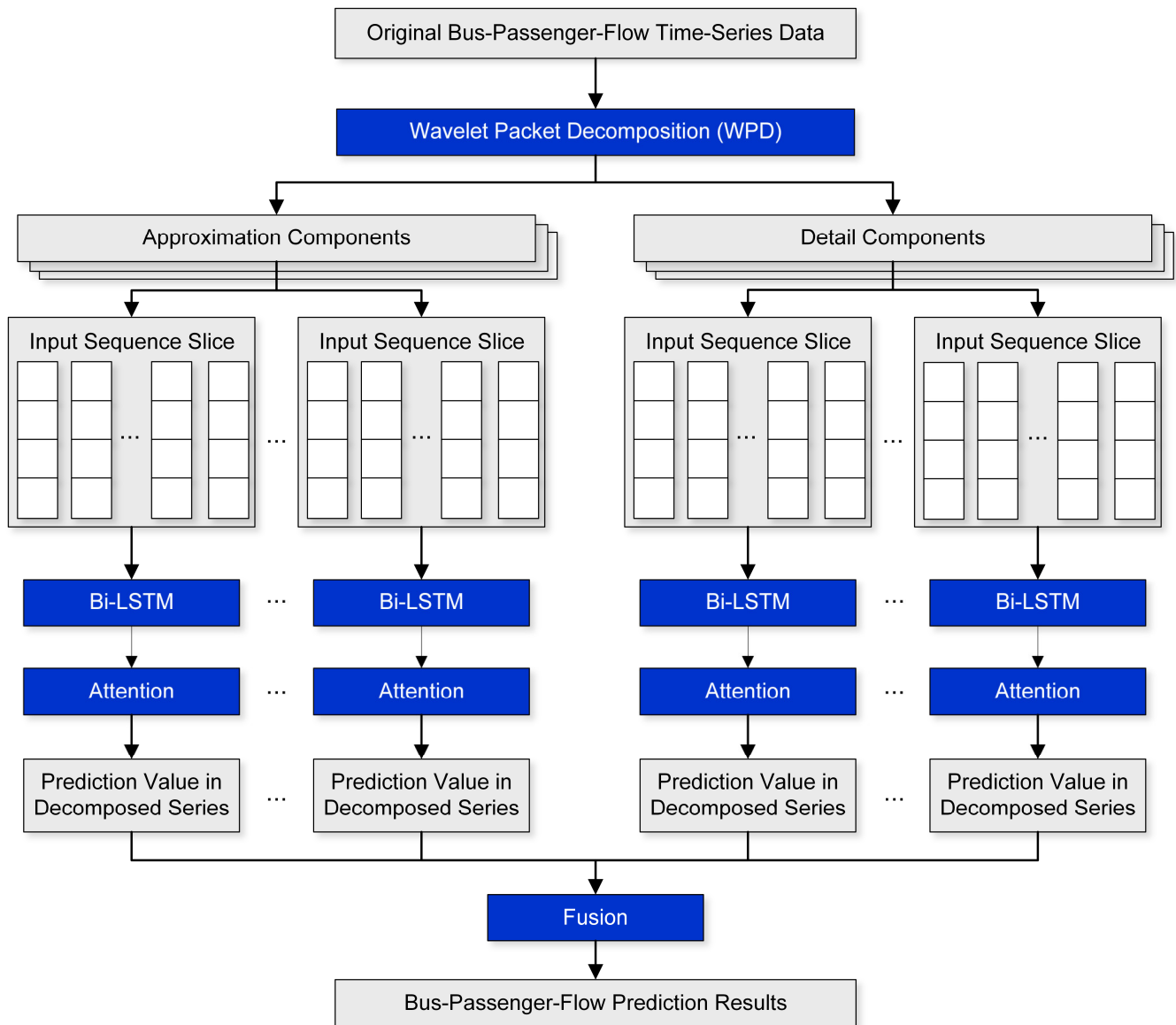

Figure 1. Comparative diagram of the 3-layer decomposition structure.

3.2. Wavelet Packet Decomposition

The wavelet method adopts low-pass and high-pass filters to decompose time series, which is an effective and widely used method to analyze time-series data in both time and frequency domains. Low-pass filter decomposition yields approximation components attributed to low frequency, while high-pass filter decomposition yields detail components attributed to high frequency [51]. WPD is a more advanced decomposition strategy based on wavelet methods developed from WT, and it allows for a more detailed decomposition of data. Whereas traditional WT can only decompose the approximation components of each layer, WPD can decompose both the approximation components and detail components. As an example, 3-layer decomposition was used to compare the structures of WT and WPD, as shown in Figure 2.
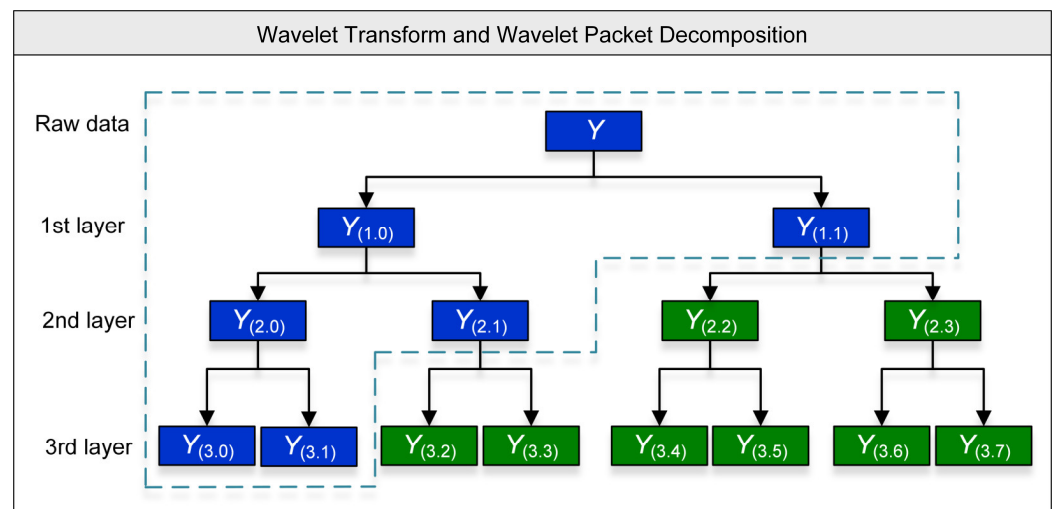
Figure 2. Comparison of the 3-layer decomposition structure of WT (dashed box) and WPD.

Compared with WT, WPD can decompose the time-series data of bus passenger flow in a more detailed way. Before loading the raw bus-passenger-flow data into the Bi-LSTM model, we used WPD to decompose the bus-passenger-flow time-series data with complex volatility into a series of different frequency components to extract the feature information of the different frequency components. The use of this feature information to train the bus-passenger-flow prediction model helped the model to better capture the changing patterns of the data trends, and it was expected to achieve more accurate passenger-flow prediction. WPD, like the WT, contains continuous and discrete transforms, which are implemented as follows:

First, determine the number of decomposition layers for wavelet packet decomposition. Different levels of decomposition of data can be performed using WPD; however, existing studies have shown that 3-level decomposition is usually used to produce the optimal accuracy of time-series prediction [50].

Next, choose the type of wavelet function; the general wavelet functions include Haar wavelet, Daubechies (dbN) wavelet, Meyer wavelet, etc.

Then, the discrete transform or continuous transform should be chosen according to the type of original data. Notably, time-series data are discrete, and discrete transform should be utilized. The continuous transform can capture more information in the original data than the discrete transform, but it can have problems of information redundancy and high computational complexity.

The original data are, finally, decomposed layer by layer according to the wavelet packet decomposition parameters set above. In Figure 2, Y represents the raw data; $Y_{(1.0)}$ and $Y_{(1.1)}$ represent the approximation component and detail component from the first layer of decomposition, respectively. WPD continues decomposing both components until the original data is mapped into 2^m wavelet subspaces, where m is the number of decomposition layers. The frequency of each subspace increases from left to right. Generally, the first 50% of the higher-frequency components are categorized as detail components, and the remaining 50% are categorized as approximation components.

3.3. Bidirectional Long–Short-Term Memory Model

The key to performing bus-passenger-flow prediction is to learn the regularity and periodicity of historical data. The time-series data components of bus passenger flow after WPD decomposition were loaded into the Bi-LSTM model. During the training process, the model learned and captured complex patterns and trends of the time-series data by simultaneously processing forward and backward information from the input series. The Bi-LSTM model was developed from the LSTM model, which was derived from RNNs' architecture and is specialized in processing time-series data. Standard RNN

models are commonly used in time-series prediction studies, though they may suffer from some problems, such as gradient disappearance, gradient explosion, and limited long-term memory capacity. The LSTM model addresses above problems through its unique memory and gate structures, allowing it to better capture the correlated characteristics inherent in time-series data [52]. This makes it a more effective solution for time-series prediction than standard RNN models. The overall structure and internal unit structure of the LSTM model are shown in Figure 3.
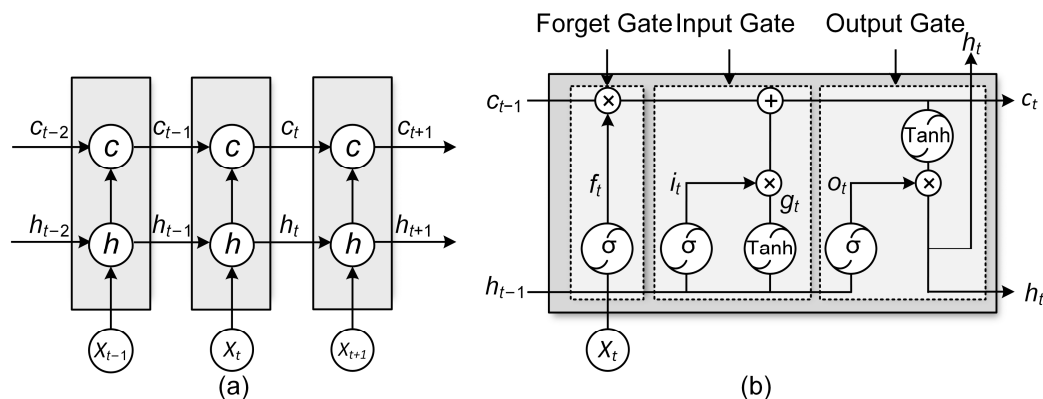

Figure 3. (**a**) Overall structure diagram of the LSTM model; (**b**) internal unit structure diagram of the LSTM model.

The LSTM network differs from traditional RNNs in that memory is rewritten at every time step. The LSTM mechanism is designed to capture and store acquired significant features into long-term memory. It then employs a selective process to either maintain, modify, or discard previously stored long-term memory, which is contingent upon the learning conditions. Through successive iterations, the neural network assigns lower weights to certain features, effectively treating them as ephemeral information that is eventually discarded from its memory. This mechanism enables the transmission of crucial characteristic information over time during iteration, thereby endowing the network with superior performance in classification tasks with long sample dependencies [53].

The LSTM model is composed of a forget gate, an input gate, and an output gate in each basic unit, as shown in Figure 3b. The forget gate at time t is determined by the input X_t, the state memory unit c_{t-1} from the previous time step, and the output state h_{t-1} from the previous time step, collectively contributing to the forgetting mechanism of the fundamental unit. The retention vector in the state memory unit is determined through the joint effect of the variation in the sigmoid and tanh functions on X_t in the input gate. The intermediate result, denoted as h_t, is determined in conjunction with the updated value of c_t and the output value of o_t. The LSTM structure formulations can be expressed as (1)–(6):

$$f_t = \sigma\left(W_{fx}X_t + W_{fh}h_{t-1} + b_f\right) \tag{1}$$

$$i_t = \sigma(W_{ix}X_t + W_{ih}h_{t-1} + b_i) \tag{2}$$

$$g_t = \phi\left(W_{gx}X_t + W_{gh}h_{t-1} + b_g\right) \tag{3}$$

$$o_t = \sigma(W_{ox}X_t + W_{oh}h_{t-1} + b_o) \tag{4}$$

$$c_t = g_t i_t + c_{t-1}f_t \tag{5}$$

$$h_t = \phi(c_t)o_t \tag{6}$$

where f_t, i_t, and o_t represent real numbers ranging from 0 to 1. Specifically, f_t denotes the proportion of forgotten long-term memory, i_t denotes the proportion of current information input into long-term memory c_t, and o_t denotes the proportion of long-term memory c_t output into the current state h_t; W_{fx}, W_{fh}, W_{ix}, W_{ih}, W_{gx}, W_{gh}, W_{ox}, and W_{oh} are the matrix weights of the multiplication of the input X_t at time t and the output state h_{t-1} at the previous time corresponding to the corresponding gate; b_f, b_i, b_g, and b_o are the bias terms of the corresponding gate; σ denotes the alteration in the sigmoid function; and ϕ denotes the change in tanh function.

The traditional LSTM model can only obtain historical information from forward to backward to predict the bus passenger flow in the future moment when dealing with bus-passenger-flow data. The information available in the time-series data of bus passenger flow after the forecast moment is not taken into account. And in the process of bus-passenger-flow time-series prediction, it is considered that a greater number of more effective features from bus-passenger-flow time-series data can be extracted by simultaneously considering the information patterns available in the data before and after the predicted moment. Therefore, in order to extract more information from the bus-passenger-flow time-series data to help improve the prediction accuracy, Bi-LSTM recurrent neural network was selected as the prediction model for bus-passenger-flow time-series data in this paper.

The Bi-LSTM model was formed via the superposition of two layers of LSTM networks. The first layer processes bus-passenger-flow time-series data from the left side, which is forward computation, to extract time-series data features. The second layer processes the bus-passenger-flow time-series data from the right side, which is the backward calculation, to obtain the time-series data features after the prediction moment. Bi-LSTM combines the information of bus-passenger-flow time-series data features from both directions. The structure of the Bi-LSTM network model is shown in Figure 4; h_t^f is the time-series data information of bus passenger flow at time t and before time t for the forward-computed LSTM network. h_t^b is the time-series data information of bus passenger flow at time t and after time t for the backward-computed LSTM network. During training, the model makes predictions in both directions simultaneously and, finally, fuses the predictions in both directions to obtain the final output, h_t.
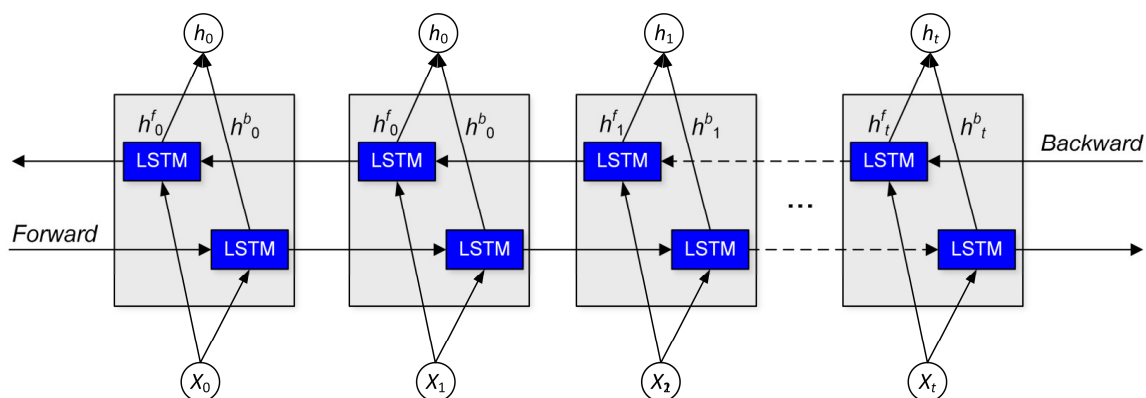

Figure 4. The structure of Bi-LSTM network model.

3.4. Attention Mechanism

After using the Bi-LSTM model to capture the features of the time-series component data of transit passenger flow that were decomposed via WPD, the hidden-layer output vector containing forward and backward data timing information can be obtained. An attention mechanism, on the other hand, can explore the impact of each data feature in the hidden layer on the predicted value of transit ridership by assigning weights to the features captured by the Bi-LSTM model according to the importance of different features.

At any given time during bus-operating hours, there are bus vehicles operating on the route, and the bus route passenger flow in each period is highly correlated with the

bus route passenger flows in the previous time periods. And both weekday and weekend bus route patronage have periodically similar fluctuations. Meanwhile, an attention mechanism can enhance the ability of the Bi-LSTM model to capture the features of transit-passenger-flow time-series data with long-term dependencies. And it helps to minimize the information loss that accompanies the network during the training process. Therefore, in this paper, an attention mechanism was combined with the Bi-LSTM model, and the attention mechanism was utilized to compute the importance of the input features of bus passenger flow at different moments. By giving different weights to the input features, the influence of the key features on bus-passenger-flow prediction was highlighted in order to improve the prediction accuracy.

The attention mechanism treated the input data as a combination of key pairs of keys and values and to assign Query to each element. Then, the similarity between Query and key was calculated to determine the value of the weight coefficient corresponding to each key, and, finally, the final weight of each feature was obtained via weighting. This included weighting and summing all the elements in the original data of length L_x and then determining the weights of the corresponding features based on Query and key. The calculation formula is as follows:

$$A(\text{Query}, \text{Source}) = \sum_{i=1}^{L_x} \text{Similarity}(\text{Query}, \text{key}_i) * \text{Value} \tag{7}$$

4. Experiment

In order to verify the effectiveness of the WPD-ATT-BiLSTM model proposed in this paper, the following experiments were conducted and are presented in this section: Using the WPD-ATT-BiLSTM model and the baseline model, single-step and multistep predictions were made for the passenger flow of bus route 363 and bus route 68 in Harbin City. Then, the prediction results of the WPD-ATT-BiLSTM model and the baseline model were visualized, and the accuracy metrics were calculated to evaluate the prediction performance of the different models.

4.1. Dataset Source and Processing

The data for the training and testing models presented in this paper came from a public transportation dataset in Harbin, China. Located in Northeast China, Harbin is the political, economic, and cultural center of the region. The bus transit system is the most important part of the urban public transportation system in Harbin City. The 500 m coverage rate of the bus stops in the main urban area of Harbin has reached 99%, providing urban residents with convenient and green bus travel services. The data were collected from the bus integrated circuit (IC) card system in Harbin, which collects the boarding information of each trip, such as the bus route, bus number, card number, card type, and boarding time. The IC card usage rate on buses is typically very high, and the current bus IC card usage rate of Harbin City has exceeded 75%. This means the IC card data capture most bus trips and passenger flows. The high coverage of the IC card data ensures their representativeness of the total bus patronage.

Two bus lines, bus 363 and bus 68, were randomly selected, and the IC card swipe data from March to October 2021 were extracted from the dataset. The swipe data obtained via extraction show a relatively high level of bus patronage on bus route 363 and a relatively low level of bus patronage on bus route 68. The data were also filtered and retained for the period of 5:30 to 19:45, when traffic is more concentrated at the station. A total of 870,000 swipes from the bus line IC card swipe records was processed using the MySQL 8.0 software, and the bus line swipe records were counted every 15 min to obtain the boarding flow of the bus line at that time. After the clustering process, the total amount of data available for the model in this paper was found to be about 30,000. To validate the effectiveness of the model proposed in this paper, the last 5 days of data in the dataset were used as a test set. And referring to other studies on deep learning training and hyperparameter tuning, the remaining 90% of the data were used as the training set, and

10% were used as the validation set. After clustering the bus card swipe records of every 15 min period, the statistical weekly passenger-flow characteristics of bus 363 and bus 68 are shown in Figures 5 and 6. In the figure, it can be seen that the bus passenger flow on weekdays has a bimodal character (morning–evening rush hours), and there is significant volatility in the data.
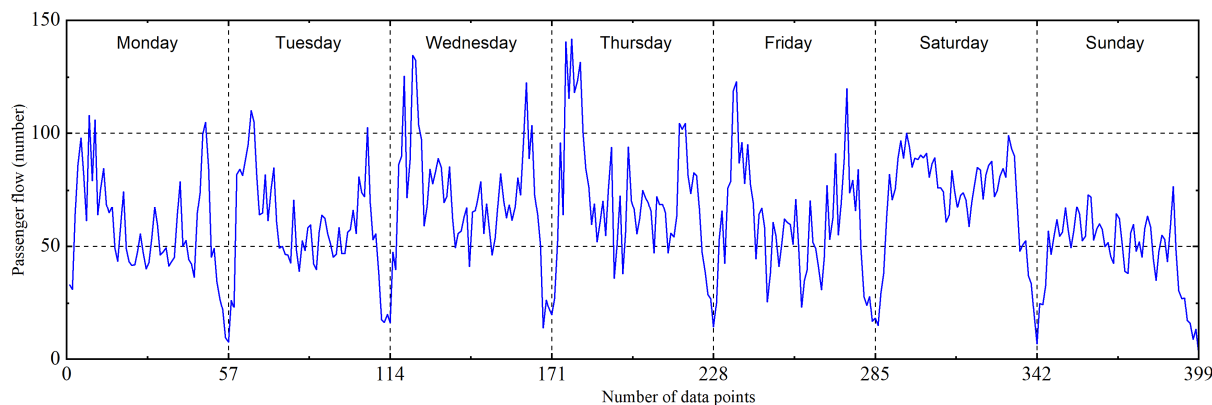

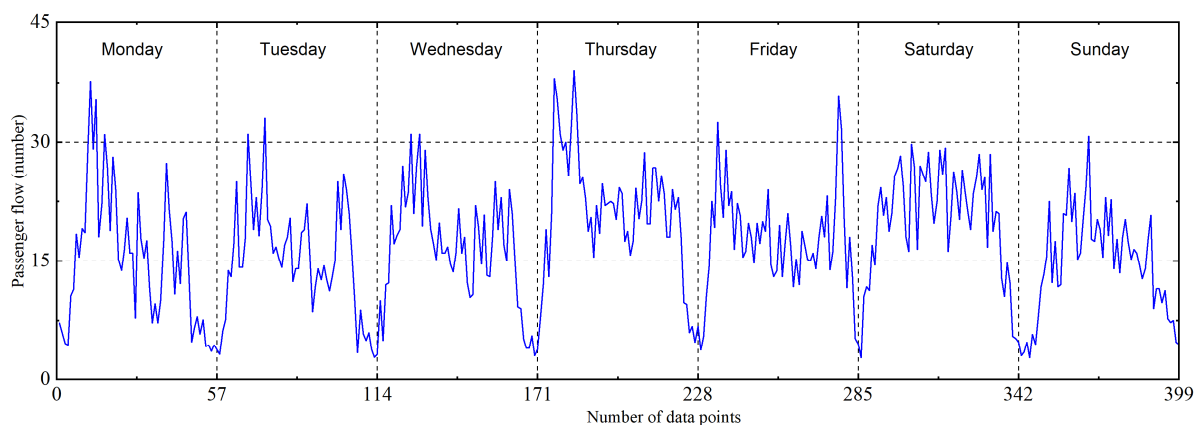**Figure 5.** Weekly passenger-flow characteristics of bus route 363.

Figure 6. Weekly passenger-flow characteristics of bus route 68.

4.2. Wavelet Packet Decomposition

It is critical to determine the number of decomposition layers and to choose the wavelet functions in the data decomposition process of WPD. Too few decomposition layers cannot fully explore the internal feature information of the original data, whereas too many decomposition layers may destroy the integrity of the original data. For bus-passenger-flow time-series data, WPD is generally employs three-layer decomposition to consider the decomposition effect and the complexity of the prediction model [50].

For wavelet function selection, the existing studies have shown that wavelet functions of the Daubechies type are used to provide a high accuracy for time-series data with periodicity [54]. Based on this, the most frequently used wavelet, Daubechies 3 (db3), was adopted as the mother wavelet in this study. The discrete wavelet transform was used to decompose the original bus-passenger-flow data of bus route 363 and bus route 68 via three-layer decomposition, and the decomposition results are shown in Figures 7 and 8.
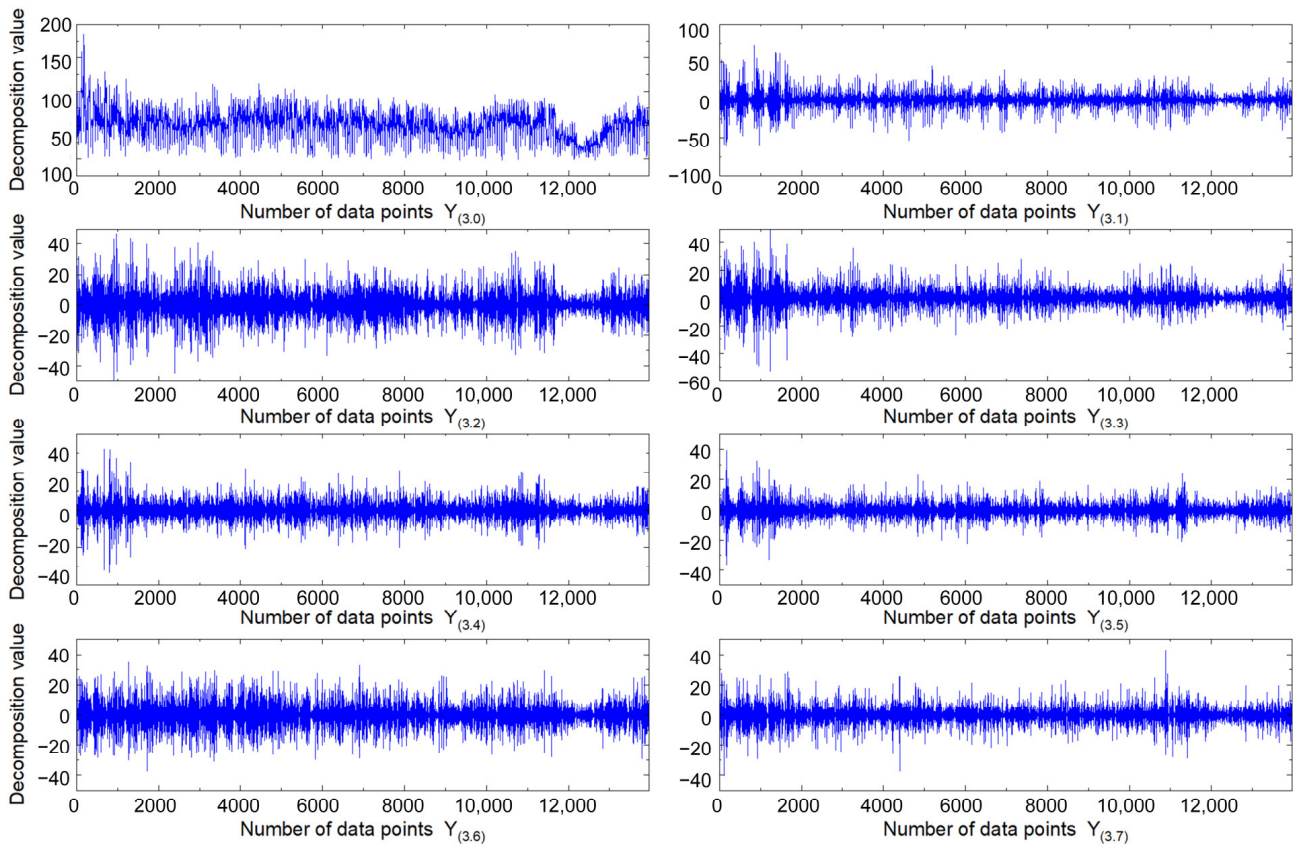
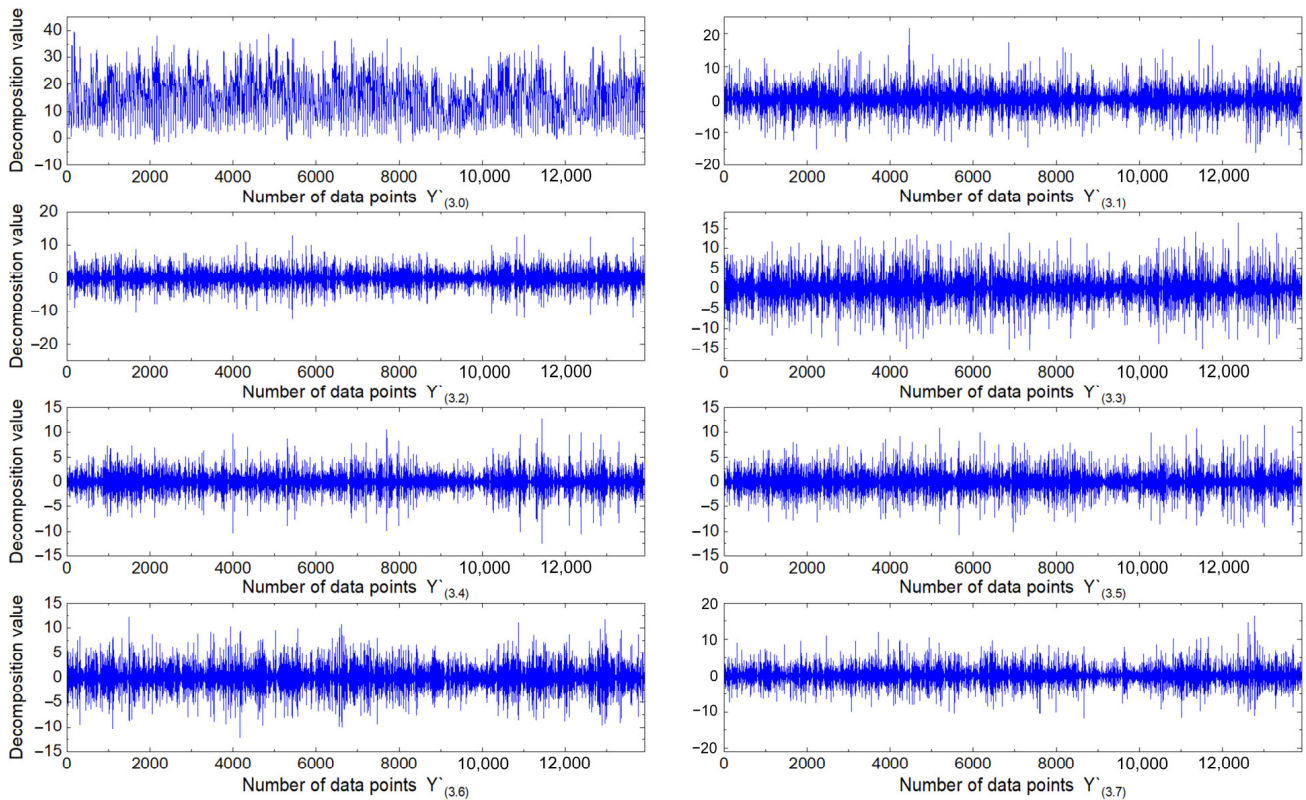**Figure 7.** Decomposition results of bus route 363's passenger flow.

Figure 8. Decomposition results of bus route 68's passenger flow.

4.3. Bi-LSTM Model Parameter Selection

The hyperparameters of Bi-LSTM were selected based on the root mean square error estimation index for the prediction accuracy. A smaller value of the index denotes a tighter match between the expected and actual values. Each experimental scenario for hyperparameter selection was run multiple times to take into consideration the variability brought on by the random initial conditions of the Bi-LSTM neural network. The determination of the hyperparameters, including the quantity of hidden units and the number of model iterations (epochs), should be achieved through a comparative analysis.

The data components after WPD processing were divided into training, validation, and testing sets, and all the data were normalized to improve the training efficiency of the model. The model presented in this paper was coded in Python 3.9 and uses Keras and TensorFlow as its deep learning frameworks. The experiments were all run on an NVIDIA RTX 3050 GPU platform. Through the parameter-tuning experiments, the number of neurons was set to 128, the epochs were set to 300 and 400, and the batch sizes were set to 16 and 32 when the model showed the optimal prediction performance. The optimizer of the model was set to Adam, and the loss function was based on the mean square error.

4.4. Precision-Estimating Indicators

In order to measure the prediction effectiveness of the passenger-flow prediction model proposed in this paper, the error of the model was evaluated using the mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE), which are calculated as shown in (8)–(10):

$$\text{MAE} = \frac{\sum\limits_{k=1}^{K} \left| y_k - \hat{y}_k \right|}{K} \tag{8}$$

$$\text{MAPE} = \frac{1}{n} \sum\limits_{k=1}^{K} \left| \frac{y_k - \hat{y}_k}{\hat{y}_k} \right| \tag{9}$$

$$\text{RMSE} = \sqrt{\frac{\sum\limits_{k=1}^{K} \left(y_k - \hat{y}_k \right)^2}{K}} \tag{10}$$

where $\hat{y}_k$ represents the prediction values, y_k represents the actual values, and K is the number of prediction values.

In addition, to compare the difference in the performance between the two prediction models, the following three percentage error indicators were also used in this study: the improvement percentages of the mean absolute error (P_{MAE}), the improvement percentages of the mean absolute percentage error (P_{MAPE}), and the improvement percentages of the root mean square error (P_{RMSE}). They are calculated as follows:

$$P_{\text{MAE}} = \left| \frac{\text{MAE}_1 - \text{MAE}_2}{\text{MAE}_1} \right| \tag{11}$$

$$P_{\text{MAPE}} = \left| \frac{\text{MAPE}_1 - \text{MAPE}_2}{\text{MAPE}_1} \right| \tag{12}$$

$$P_{\text{RMSE}} = \left| \frac{\text{RMSE}_1 - \text{RMSE}_2}{\text{RMSE}_1} \right| \tag{13}$$

4.5. Model Prediction Results and Comparison

In this section, the experimental data are visualized, and the results of the prediction experiments are shown in detail. Aiming to verify the effectiveness of the WPD-ATT-BiLSTM model proposed in this paper, the XGBoost model, SVR model, LSTM model,

Bi-LSTM model, Bi-LSTM with attention mechanism (ATT-BiLSTM) model, and Bi-LSTM with wavelet packet decomposition (WPD-BiLSTM) model were used as the benchmark models, and the accuracy estimation metrics of the prediction results of each model were calculated.

After loading a series of decomposition data into the prediction model to obtain the predicted value of each decomposition sequence, the predicted value of each sequence was fused to obtain the final bus-flow prediction results. The formula for fusing the decomposition components is shown below:

$$y = \sum_{n=0}^{2^m-1} y_{(m,n)} \tag{14}$$

where y is the final prediction, $y_{(m,n)}$ is the prediction value of the different frequency components, m is the number of decomposition layers, and n is the amount of data decomposition.

Figure 5 shows scatter density plots of the predicted transit ridership. The scatter density plots combine scatter plots and kernel density estimation plots to show the distribution of the data points and to help us visualize and compare the predictive performance of each model with respect to the bus patronage on bus route 363. The straight line in the graphs represents the true values, and the dots represent the predicted values. The more concentrated the points are near a straight line, the better the predictive performance of the model. In addition, the color of the scatter density plots also reflects the density distribution of the data points. The RMSE values corresponding to the predictions of each prediction model are also labeled in the upper left corner of the figure.

As can be seen in Figure 9, the WPD-ATT-BiLSTM model has the best prediction performance. The reasons for this are that the points of the scatter density plot of the model are most concentrated around the straight line and that the model has the smallest RMSE value for its predicted values.
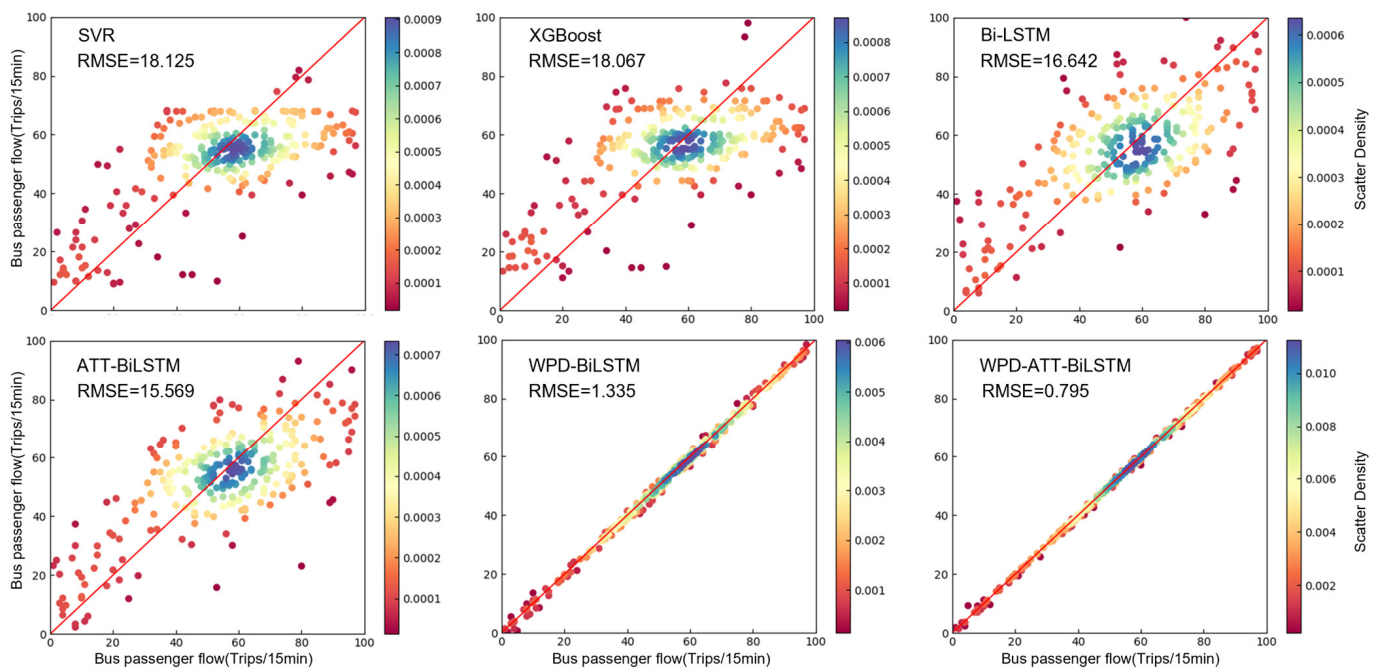

Figure 9. Scattered density plots of predicted values of public transportation patronage.

In order to further validate the effectiveness of the WPD-ATT-BiLSTM model proposed in this paper with respect to prediction, on the basis of adding the multistep prediction

experiments of the bus passenger flow, the bus-passenger-flow dataset of bus route 68 was also added for model training and testing.

The multistep prediction of the bus passenger flow over multiple time spans allows for a wider range of applications in prediction. To assess the multistep predictive capability of the model proposed in this study, the variability of the prediction accuracy was examined over three different time horizons: 15 min (one-step horizon), 30 min (two-step horizon), and 45 min (three-step horizon). In addition, the single-step and multistep prediction results were compared and analyzed with those of the XGBoost model, SVR model, Bi-LSTM model, ATT-BiLSTM model, and WPD-BiLSTM model.

Figures 10 and 11 illustrate the predictive performance of the various models on the test set over the one-step, two-step, and three-step horizons, respectively. Furthermore, the calculation results of the predictive precision indexes of the WPD-ATT-BiLSTM model and the other prediction models that utilized the same dataset are shown in Tables 1 and 2.
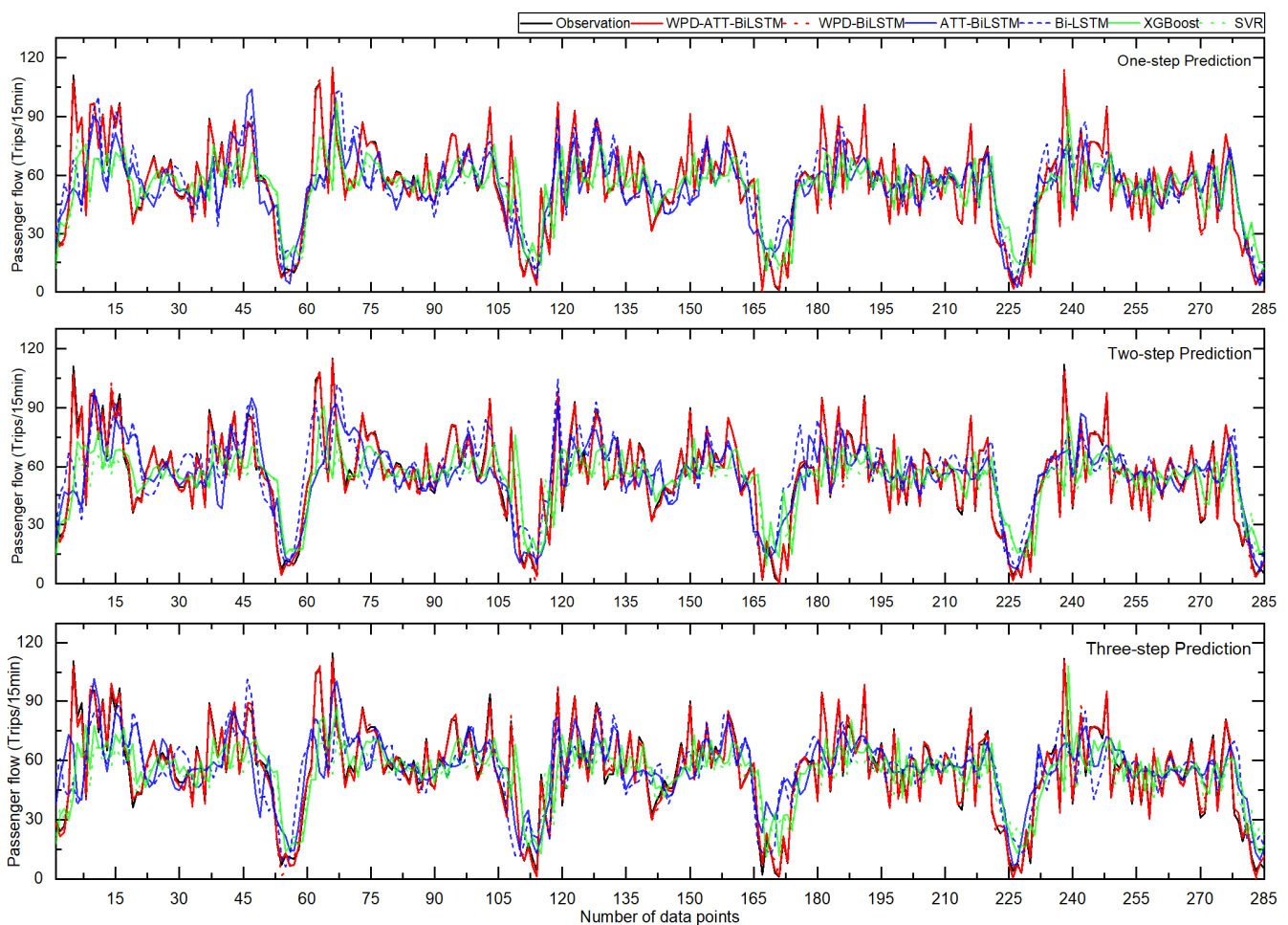

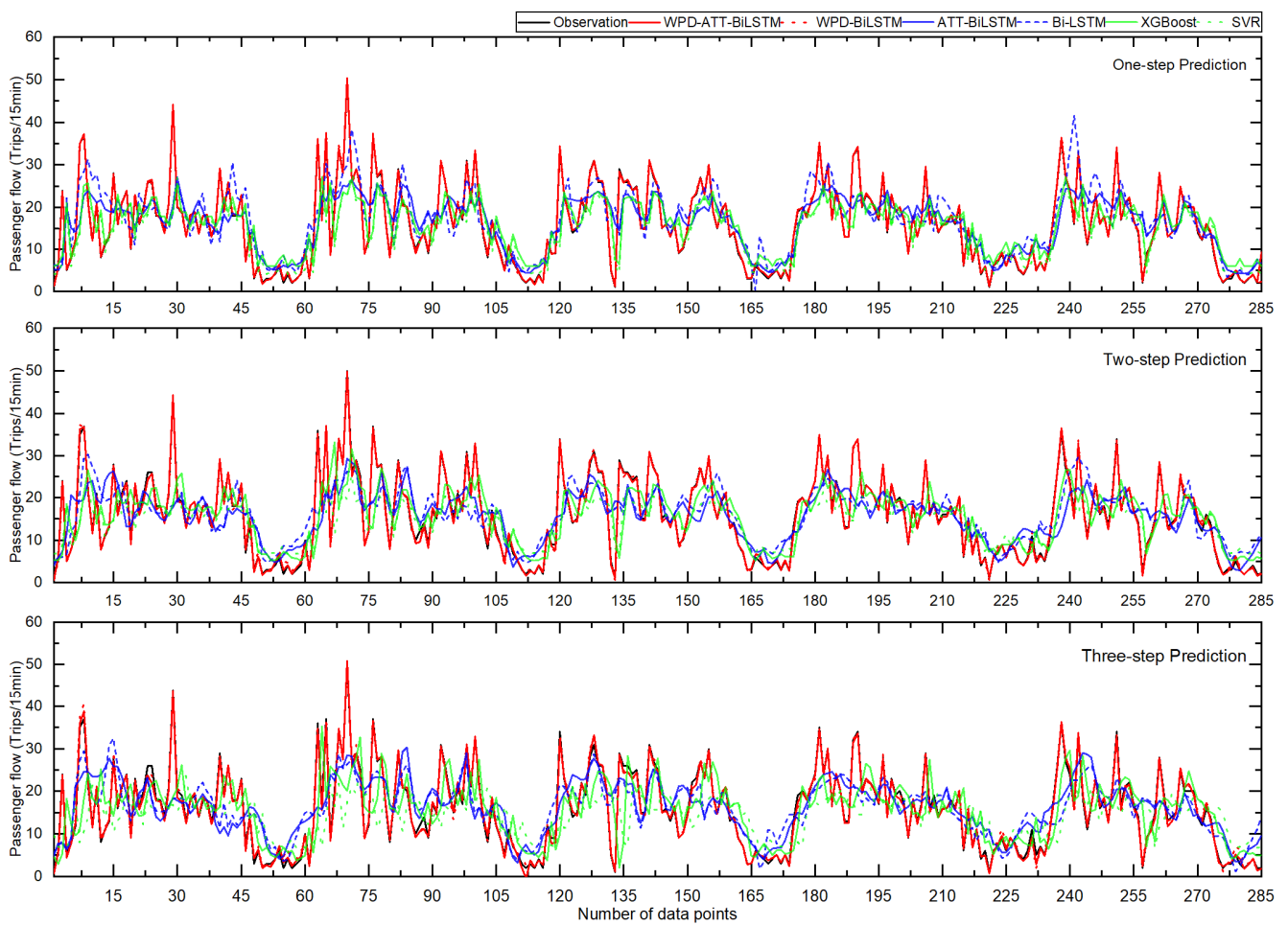**Figure 10.** Bus route 363 passenger-flow multistep prediction results.

Figure 11. Bus route 68 passenger-flow multistep prediction results.

Table 1. Precision estimation indicator values of multistep prediction results of each model for bus passenger flow on bus route 363.

Prediction Models	MAE			MAPE (%)			RMSE		
	1 Step	2 Steps	3 Steps	1 Step	2 Steps	3 Steps	1 Step	2 Steps	3 Steps
WPD-ATT-BiLSTM	0.582	1.162	1.670	2.283	4.621	8.547	0.795	1.586	2.102
WPD-BiLSTM	1.046	1.325	1.723	7.951	7.231	9.690	1.335	1.882	2.352
ATT-BiLSTM	12.111	12.545	13.219	25.387	24.303	26.031	15.569	15.956	16.686
Bi-LSTM	13.001	13.324	13.593	25.996	26.070	28.623	16.642	16.748	17.474
XGBOOST	14.070	14.047	14.062	30.749	30.780	29.659	18.067	17.944	17.917
SVR	13.927	14.165	14.375	33.197	31.279	30.943	18.125	18.270	18.368

Table 2. Precision estimation indicator values of multistep prediction results of each model for bus passenger flow on bus route 68.

Prediction Models	MAE			MAPE (%)			RMSE		
	1 Step	2 Steps	3 Steps	1 Step	2 Steps	3 Steps	1 Step	2 Steps	3 Steps
WPD-ATT-BiLSTM	0.235	0.376	0.584	2.740	4.130	4.357	0.275	0.507	0.792
WPD-BiLSTM	0.266	0.412	0.617	2.982	4.796	6.840	0.331	0.590	0.885
ATT-BiLSTM	5.026	5.276	5.370	33.346	36.472	35.578	6.786	7.016	7.115
Bi-LSTM	5.136	5.451	5.506	38.047	36.662	38.319	6.826	7.057	7.123
XGBOOST	5.341	5.483	5.642	37.926	39.587	45.041	7.168	7.369	7.742
SVR	5.264	5.473	6.511	41.709	40.602	50.234	7.266	7.394	8.629

In order to conduct a thorough comparison of the predictive capabilities of the model presented in this paper and the other benchmark models, three percentage error indicators were computed. These criteria serve to effectively measure the variance in performance between the two prediction models, namely, P_{MAE}, P_{MAPE}, and P_{RMSE}, respectively. The results are shown in Tables 3 and 4.

Table 3. Percentage performance improvement indicators of the WPD-ATT-BiLSTM model compared to the other models in predicting bus passenger flow on bus route 363.

Comparison Models	P_{MAE} (%)			P_{MAPE} (%)			P_{RMSE} (%)		
	1 Step	2 Steps	3 Steps	1 Step	2 Steps	3 Steps	1 Step	2 Steps	3 Steps
WPD-BiLSTM	44.359	12.302	3.076	71.287	36.095	11.796	40.449	15.728	10.629
ATT-BiLSTM	95.194	90.737	87.367	91.007	80.986	67.166	94.894	90.060	87.403
Bi-LSTM	95.523	91.279	87.714	91.218	82.275	70.139	95.223	90.530	87.971
XGBOOST	95.864	91.728	88.124	92.575	84.987	71.182	95.600	91.161	88.268
SVR	95.821	91.797	88.383	93.123	85.227	72.378	95.614	91.319	88.556

Table 4. Percentage performance improvement indicators of the WPD-ATT-BiLSTM model compared to the other models in predicting bus passenger flow on bus route 68.

Comparison Models	P_{MAE} (%)			P_{MAPE} (%)			P_{RMSE} (%)		
	1 Step	2 Steps	3 Steps	1 Step	2 Steps	3 Steps	1 Step	2 Steps	3 Steps
WPD-BiLSTM	11.654	8.738	5.348	8.115	13.887	36.301	16.918	14.068	10.508
ATT-BiLSTM	95.324	92.873	89.378	91.783	88.676	87.761	95.948	92.774	88.983
Bi-LSTM	95.424	93.102	89.496	92.798	88.735	87.716	95.971	92.816	89.113
XGBOOST	95.600	93.142	89.649	92.775	89.567	90.327	96.164	93.120	89.770
SVR	95.536	93.130	91.031	93.431	89.828	91.327	96.215	93.143	90.822

5. Discussion

Based on the prediction experiment results in the previous section, it can be seen that the WPD-ATT-BiLSTM model proposed in this paper exhibits a satisfactory prediction performance for the different passenger-flow datasets of the bus routes. This indicates that the model does not only perform well on the passenger-flow data of a particular route but can also cope well with the problem of predicting the time-series data of the passenger flow of different bus routes. As can be seen in Tables 1 and 2, among all the prediction models, the proposed model has the smallest values for the estimated indexes of the prediction accuracy over a single step and multiple steps, which proves the effectiveness of the model. Figures 5 and 6 also visualize that the model's fit of the predicted values to the actual values is optimal.

The following is specifically shown in Tables 1–4:

Among the models involved in the prediction experiment, the WPD-ATT-BILSTM model achieved the best predictive performance for one-step to three-step prediction on the test set, which demonstrates the effectiveness of the hybrid prediction model for the bus passenger flow proposed in this paper in improving the prediction accuracy.

The hybrid-strategy prediction model, WPD-ATT-BiLSTM, proposed in this paper, obtained a better prediction result accuracy than those of the machine learning models (SVR and XGBoost) and the deep learning model (Bi-LSTM) in the experiments. Taking the prediction result data of the one-step prediction of route 363's bus passenger flow as an example, in comparison with SVR and XGBoost, the model improved the prediction performance of the MAE indicator by 95.821% and 95.864%, the prediction performance of the MAPE indicator by 92.575% and 93.123%, and the prediction performance of the RMSE indicator by 92.575% and 93.123%. It can be seen that the bus-passenger-flow prediction model proposed in this paper, which combines a decomposition strategy and an attention mechanism, has an excellent prediction performance that is significantly better than those of the traditional machine learning models and the deep learning model. This indicates

that the mixed-strategy prediction model proposed in this paper, which combines WPD, an attention mechanism, and the Bi-LSTM model, fully utilizes the advantages of the decomposition strategy, the attention mechanism, and the deep learning model to better and adequately capture the features of the time series, which leads to a great improvement in the prediction performance.

WPD-ATT-BiLSTM and ATT-BiLSTM are the prediction models with the addition of an attention mechanism, and WPD-BiLSTM and BiLSTM are the prediction models without an attention mechanism; by comparing the prediction results of these two groups, the positive effect of the attention mechanism on the improvement in the prediction performance was verified. Taking the prediction result data of the one-step prediction of route 363's bus passenger flow as an example, the WPD-ATT-BiLSTM model and ATT-BiLSTM model improved the prediction performance of the MAE indicator by 44.359% and 6.846% over the WPD-BiLSTM model and the BiLSTM model; the prediction performance of the MAPE indicator improved by 71.287% and 2.343%, and the prediction performance of the RMSE metrics improved by 40.449% and 6.448%.

WPD-ATT-BiLSTM and WPD-BiLSTM are the prediction models using the WPD decomposition strategy, and ATT-BiLSTM and Bi-LSTM are the prediction models without the decomposition strategy; the positive effect of the decomposition strategy on the prediction performance enhancement's was verified by comparing these two groups of models. Still taking the prediction result data of the one-step prediction of the bus passenger flow on route 363 as an example, the WPD-ATT-BiLSTM model and the WPD-BiLSTM model improved the prediction performance of the MAE metrics by 95.194% and 91.954% over the ATT-BiLSTM model and the Bi-LSTM model; the prediction performance of the MAPE metrics improved by 91.007% and 69.415%, and the prediction performance of the RMSE metrics improved by 94.894% and 91.978%.

6. Conclusions

Extracting the key features in time-series data and capturing the changing patterns of the data are necessary for accurate bus-passenger-flow prediction. However, affected by various factors, bus-passenger-flow data have obvious volatility and nonlinearity. It is challenging to directly extract features from the original bus-passenger-flow data with complex information, resulting in the unsatisfactory prediction accuracy of many models. In this paper, a new mixed-strategy prediction model, the WPD-ATT-BiLSTM model, is proposed. The model is based on the deep learning Bi-LSTM model, and it addresses the problem of the difficulty for a single deep learning model to fully capture the characteristics of bus-passenger-flow time-series data. To this end, it innovatively introduces a decomposition strategy and an attention mechanism to improve the ability of the deep learning model to fit complex time-series data.

The Bi-LSTM model was employed to learn the time-series data features of the bus passenger flow from both directions. Additionally, the WPD data decomposition method was introduced to convert the raw and fluctuating bus-passenger-flow data into a series of smoother data components. This addresses the issue that models may struggle to fully capture the features of nonsmooth raw data during the training process. Moreover, the model also incorporates an attention mechanism, which assigns weights to the input features based on their importance. This further enhances the data feature extraction ability of the model and reduces the interference of irrelevant information.

The model proposed in this paper demonstrated a satisfactory prediction performance in the experiments, validating its effectiveness for bus-passenger-flow prediction. This model can be applied to the practical dynamic prediction of the bus passenger flow. It provides an effective reference foundation for the dynamic scheduling of buses while also assisting bus operating companies in the rational allocation of their capacity resources and in reducing their operating costs. The main findings of this paper are as follows:

- Compared with machine learning models (SVR and XGBoost) and deep learning models (Bi-LSTM), the WPD-ATT-BILSTM model proposed in this paper has a significant

advantage in extracting the temporal features of the bus passenger flow. On the basis of the Bi-LSTM model's bidirectional learning of the bus-passenger-flow time-series data features, the model's prediction performance was greatly improved by adding a data decomposition layer and an attention mechanism layer.

- It was demonstrated that weighting the input features using an attention mechanism can have a positive effect on the model prediction performance's improvement by comparing the WPD-ATT-BiLSTM and ATT-BiLSTM prediction models with the addition of an attention mechanism to the WPD-BiLSTM and BiLSTM prediction models without an attention mechanism. The WPD-ATT-BiLSTM and WPD-BiLSTM prediction models with the addition of the WPD decomposition strategy were compared with the ATT-BiLSTM and Bi-LSTM models without the decomposition strategy, and it was demonstrated that the use of the decomposition strategy to convert the fluctuating raw data into smoother data components can effectively improve the prediction performance of the models.
- Single-step and multistep prediction experiments of the model were carried out using the passenger-flow data of two bus routes in Harbin City; the experimental data were visualized, the prediction accuracy assessment index was calculated, and the results showed that the prediction performance of the WPD-ATT-BiLSTM model proposed in this paper is the optimal and most stable prediction performance compared to those of the benchmark models.

In summary, this paper proposes a new bus-passenger-flow prediction model based on WPD, an attention mechanism, and Bi-LSTM, which provide strong support for buses' dynamic scheduling. Thus, it can better satisfy the passengers' demands for bus travel, and the attraction of buses can be improved. The bus passenger flow has complex fluctuation characteristics and is affected by numerous factors. In future research, in addition to exploring the time-series correlation of the bus passenger flow, the correlation of the bus passenger flow over a spatial distribution should also be considered. Therefore, there is a scope to further improve the predictive accuracy and effectiveness of the model.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all the subjects involved in this study.

Data Availability Statement: Not applicable.

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

References

1. Javid, M.A.; Ali, N.; Hussain Shah, S.A.; Abdullah, M. Travelers' Attitudes Toward Mobile Application-Based Public Transport Services in Lahore. *Sage Open* **2021**, *11*, 2158244020988709. [CrossRef]
2. Leong, W.; Goh, K.; Hess, S.; Murphy, P. Improving Bus Service Reliability: The Singapore Experience. *Res. Transp. Econ.* **2016**, *59*, 40–49. [CrossRef]
3. Tang, T.; Fonzone, A.; Liu, R.; Choudhury, C. Multi-Stage Deep Learning Approaches to Predict Boarding Behaviour of Bus Passengers. *Sustain. Cities Soc.* **2021**, *73*, 103111. [CrossRef]
4. Wu, W.; Xia, Y.; Jin, W. Predicting Bus Passenger Flow and Prioritizing Influential Factors Using Multi-Source Data: Scaled Stacking Gradient Boosting Decision Trees. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 2510–2523. [CrossRef]

5. Liu, Y.; Lyu, C.; Liu, X.; Liu, Z. Automatic Feature Engineering for Bus Passenger Flow Prediction Based on Modular Convolutional Neural Network. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 2349–2358. [CrossRef]

6. Zhao, T.; Huang, Z.; Tu, W.; He, B.; Cao, R.; Cao, J.; Li, M. Coupling Graph Deep Learning and Spatial-Temporal Influence of Built Environment for Short-Term Bus Travel Demand Prediction. *Comput. Environ. Urban Syst.* **2022**, *94*, 101776. [CrossRef]

7. Williams, B.M.; Hoel, L.A. Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results. *J. Transp. Eng.* **2003**, *129*, 664–672. [CrossRef]

8. Stathopoulos, A.; Karlaftis, M.G. A Multivariate State Space Approach for Urban Traffic Flow Modeling and Prediction. *Transp. Res. Part C Emerg. Technol.* **2003**, *11*, 121–135. [CrossRef]

9. Castillo, E.; Maria Menendez, J.; Sanchez-Cambronero, S. Predicting Traffic Flow Using Bayesian Networks. *Transp. Res. Part B Methodol.* **2008**, *42*, 482–509. [CrossRef]

10. Sapankevych, N.L.; Sankar, R. Time Series Prediction Using Support Vector Machines: A Survey. *IEEE Comput. Intell. Mag.* **2009**, *4*, 24–38. [CrossRef]

11. Vlahogianni, E.I.; Karlaftis, M.G.; Golias, J.C. Optimized and Meta-Optimized Neural Networks for Short-Term Traffic Flow Prediction: A Genetic Approach. *Transp. Res. Part C Emerg. Technol.* **2005**, *13*, 211–234. [CrossRef]

12. Bharti; Redhu, P.; Kumar, K. Short-Term Traffic Flow Prediction Based on Optimized Deep Learning Neural Network: PSO-Bi-LSTM. *Phys. Stat. Mech. Its Appl.* **2023**, *625*, 129001. [CrossRef]

13. Chen, T.; Fang, J.; Xu, M.; Tong, Y.; Chen, W. Prediction of Public Bus Passenger Flow Using Spatial-Temporal Hybrid Model of Deep Learning. *J. Transp. Eng. Part Syst.* **2022**, *148*, 04022007. [CrossRef]

14. Kashi, S.O.M.; Akbarzadeh, M. A Framework for Short-Term Traffic Flow Forecasting Using the Combination of Wavelet Transformation and Artificial Neural Networks. *J. Intell. Transp. Syst.* **2019**, *23*, 60–71. [CrossRef]

15. Dunne, S.; Ghosh, B. Weather Adaptive Traffic Prediction Using Neurowavelet Models. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 370–379. [CrossRef]

16. Barthelemy, J.; Verstaevel, N.; Forehead, H.; Perez, P. Edge-Computing Video Analytics for Real-Time Traffic Monitoring in a Smart City. *Sensors* **2019**, *19*, 2048. [CrossRef]

17. Zheng, Z.; Shi, L.; Sun, L.; Du, J. Short-Term Traffic Flow Prediction Based on Sparse Regression and Spatio-Temporal Data Fusion. *IEEE Access* **2020**, *8*, 142111–142119. [CrossRef]

18. Cai, L.; Zhang, Z.; Yang, J.; Yu, Y.; Zhou, T.; Qin, J. A Noise-Immune Kalman Filter for Short-Term Traffic Flow Forecasting. *Phys.-Stat. Mech. Its Appl.* **2019**, *536*, 122601. [CrossRef]

19. Lu, J.; Peng, J.; Chen, J.; Sugeng, K.A. Prediction Method of Autoregressive Moving Average Models for Uncertain Time Series. *Int. J. Gen. Syst.* **2020**, *49*, 546–572. [CrossRef]

20. Shahriari, S.; Ghasri, M.; Sisson, S.A.; Rashidi, T. Ensemble of ARIMA: Combining Parametric and Bootstrapping Technique for Traffic Flow Prediction. *Transp. A Transp. Sci.* **2020**, *16*, 1552–1573. [CrossRef]

21. Sun, B.; Sun, T.; Zhang, Y.; Jiao, P. Urban Traffic Flow Online Prediction Based on Multi-Component Attention Mechanism. *IET Intell. Transp. Syst.* **2020**, *14*, 1249–1258. [CrossRef]

22. Zhang, W.; Zhu, K.; Zhang, S.; Chen, Q.; Xu, J. Dynamic Graph Convolutional Networks Based on Spatiotemporal Data Embedding for Traffic Flow Forecasting. *Knowl. Based Syst.* **2022**, *250*, 109028. [CrossRef]

23. Chen, X.; Wu, S.; Shi, C.; Huang, Y.; Yang, Y.; Ke, R.; Zhao, J. Sensing Data Supported Traffic Flow Prediction via Denoising Schemes and ANN: A Comparison. *IEEE Sens. J.* **2020**, *20*, 14317–14328. [CrossRef]

24. Wang, D.; Wang, C.; Xiao, J.; Xiao, Z.; Chen, W.; Havyarimana, V. Bayesian Optimization of Support Vector Machine for Regression Prediction of Short-Term Traffic Flow. *Intell. Data Anal.* **2019**, *23*, 481–497. [CrossRef]

25. Liu, L.; Chen, R.-C.; Zhao, Q.; Zhu, S. Applying a Multistage of Input Feature Combination to Random Forest for Improving MRT Passenger Flow Prediction. *J. Ambient Intell. Humaniz. Comput.* **2019**, *10*, 4515–4532. [CrossRef]

26. Sun, B.; Sun, T.; Jiao, P. Spatio-Temporal Segmented Traffic Flow Prediction with ANPRS Data Based on Improved XGBoost. *J. Adv. Transp.* **2021**, *2021*, 5559562. [CrossRef]

27. Lu, X.; Chen, C.; Gao, R.; Xing, Z. Prediction of High-Speed Traffic Flow around City Based on BO-XGBoost Model. *Symmetry* **2023**, *15*, 1453. [CrossRef]

28. Huang, P.; Wen, C.; Fu, L.; Peng, Q.; Tang, Y. A Deep Learning Approach for Multi-Attribute Data: A Study of Train Delay Prediction in Railway Systems. *Inf. Sci.* **2020**, *516*, 234–253. [CrossRef]

29. Nguyen, H.; Kieu, L.-M.; Wen, T.; Cai, C. Deep Learning Methods in Transportation Domain: A Review. *IET Intell. Transp. Syst.* **2018**, *12*, 998–1004. [CrossRef]

30. Lin, Y.; Zhang, J.; Liu, H. Deep Learning Based Short-Term Air Traffic Flow Prediction Considering Temporal-Spatial Correlation. *Aerosp. Sci. Technol.* **2019**, *93*, 105113. [CrossRef]

31. Chen, C.; Liu, Z.; Wan, S.; Luan, J.; Pei, Q. Traffic Flow Prediction Based on Deep Learning in Internet of Vehicles. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 3776–3789. [CrossRef]

32. Nagaraj, N.; Gururaj, H.L.; Swathi, B.H.; Hu, Y.-C. Passenger Flow Prediction in Bus Transportation System Using Deep Learning. *Multimed. Tools Appl.* **2022**, *81*, 12519–12542. [CrossRef] [PubMed]

33. Du, B.; Peng, H.; Wang, S.; Bhuiyan, M.Z.A.; Wang, L.; Gong, Q.; Liu, L.; Li, J. Deep Irregular Convolutional Residual LSTM for Urban Traffic Passenger Flows Prediction. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 972–985. [CrossRef]

34. Han, Y.; Wang, C.; Ren, Y.; Wang, S.; Zheng, H.; Chen, G. Short-Term Prediction of Bus Passenger Flow Based on a Hybrid Optimized LSTM Network. *ISPRS Int. J. Geo. Inf.* **2019**, *8*, 366. [CrossRef]

35. He, P.; Jiang, G.; Lam, S.-K.; Sun, Y. Learning Heterogeneous Traffic Patterns for Travel Time Prediction of Bus Journeys. *Inf. Sci.* **2020**, *512*, 1394–1406. [CrossRef]

36. Abduljabbar, R.L.; Dia, H.; Tsai, P.-W. Unidirectional and Bidirectional LSTM Models for Short-Term Traffic Prediction. *J. Adv. Transp.* **2021**, *2021*, 5589075. [CrossRef]

37. Zhai, Y.; Wan, Y.; Wang, X. Optimization of Traffic Congestion Management in Smart Cities under Bidirectional Long and Short-Term Memory Model. *J. Adv. Transp.* **2022**, *2022*, 3305400. [CrossRef]

38. Li, Z.; Xu, H.; Gao, X.; Wang, Z.; Xu, W. Fusion Attention Mechanism Bidirectional LSTM for Short-Term Traffic Flow Prediction. *J. Intell. Transp. Syst.* **2022**, 1–14. [CrossRef]

39. Ma, Z.; Xing, J.; Mesbah, M.; Ferreira, L. Predicting Short-Term Bus Passenger Demand Using a Pattern Hybrid Approach. *Transp. Res. Part C Emerg. Technol.* **2014**, *39*, 148–163. [CrossRef]

40. Glisovic, N.; Milenkovic, M.; Bojovic, N.; Svadlenka, L.; Avramovic, Z. A Hybrid Model for Forecasting the Volume of Passenger Flows on Serbian Railways. *Oper. Res.* **2016**, *16*, 271–285. [CrossRef]

41. Xu, D.; Wang, Y.; Jia, L.; Qin, Y.; Dong, H. Real-Time Road Traffic State Prediction Based on ARIMA and Kalman Filter. *Front. Inf. Technol. Electron. Eng.* **2017**, *18*, 287–302. [CrossRef]

42. Lin, X.; Huang, Y. Short-Term High-Speed Traffic Flow Prediction Based on ARIMA-GARCH-M Model. *Wirel. Pers. Commun.* **2021**, *117*, 3421–3430. [CrossRef]

43. Li, Y.; Ma, C. Short-Time Bus Route Passenger Flow Prediction Based on a Secondary Decomposition Integration Method. *J. Transp. Eng. Part A Syst.* **2023**, *149*, 04022132. [CrossRef]

44. Yang, B.; Sun, S.; Li, J.; Lin, X.; Tian, Y. Traffic Flow Prediction Using LSTM with Feature Enhancement. *Neurocomputing* **2019**, *332*, 320–327. [CrossRef]

45. Abdelraouf, A.; Abdel-Aty, M.; Yuan, J. Utilizing Attention-Based Multi-Encoder-Decoder Neural Networks for Freeway Traffic Speed Prediction. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 11960–11969. [CrossRef]

46. Qian, Z.; Pei, Y.; Zareipour, H.; Chen, N. A Review and Discussion of Decomposition-Based Hybrid Models for Wind Energy Forecasting Applications. *Appl. Energy* **2019**, *235*, 939–953. [CrossRef]

47. Khandelwal, I.; Adhikari, R.; Verma, G. Time Series Forecasting Using Hybrid ARIMA and ANN Models Based on DWT Decomposition. *Procedia Comput. Sci.* **2015**, *48*, 173–179. [CrossRef]

48. Diao, Z.; Zhang, D.; Wang, X.; Xie, K.; He, S.; Lu, X.; Li, Y. A Hybrid Model for Short-Term Traffic Volume Prediction In Massive Transportation Systems. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 935–946. [CrossRef]

49. Zhu, S.; Zhao, Y.; Zhang, Y.; Li, Q.; Wang, W.; Yang, S. Short-Term Traffic Flow Prediction with Wavelet and Multi-Dimensional Taylor Network Model. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 3203–3208. [CrossRef]

50. El-Hendawi, M.; Wang, Z. An Ensemble Method of Full Wavelet Packet Transform and Neural Network for Short Term Electrical Load Forecasting. *Electr. Power Syst. Res.* **2020**, *182*, 106265. [CrossRef]

51. Aghajani, A.; Kazemzadeh, R.; Ebrahimi, A. A Novel Hybrid Approach for Predicting Wind Farm Power Production Based on Wavelet Transform, Hybrid Neural Networks and Imperialist Competitive Algorithm. *Energy Convers. Manag.* **2016**, *121*, 232–240. [CrossRef]

52. Liu, X.; Lin, Z.; Feng, Z. Short-Term Offshore Wind Speed Forecast by Seasonal ARIMA—A Comparison against GRU and LSTM. *Energy* **2021**, *227*, 120492. [CrossRef]

53. Wu, J.; Wang, Z. A Hybrid Model for Water Quality Prediction Based on an Artificial Neural Network, Wavelet Transform, and Long Short-Term Memory. *Water* **2022**, *14*, 610. [CrossRef]

54. Tang, J.; Chen, X.; Hu, Z.; Zong, F.; Han, C.; Li, L. Traffic Flow Prediction Based on Combination of Support Vector Machine and Data Denoising Schemes. *Phys. Stat. Mech. Its Appl.* **2019**, *534*, 120642. [CrossRef]