Article

A Spatiotemporal Locomotive Axle Temperature Prediction Approach Based on Ensemble Graph Convolutional Recurrent Unit Networks

Ye Li [1] , Limin Yang [2], Yutong Wan [2] and Yu Bai [3],*

1   Key Laboratory of Traffic Safety on Track of Ministry of Education, School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China
2   Unit 93147 of the Chinese People's Liberation Army, China
3   School of Information and Engineering, Hebei University of Science and Technology, Shijiazhuang 050001, China
*   Correspondence: baiyu0584@gmail.com

Abstract: Spatiotemporal axle temperature forecasting is crucial for real-time failure detection in locomotive control systems, significantly enhancing reliability and facilitating early maintenance. Motivated by the need for more accurate and reliable prediction models, this paper proposes a novel ensemble graph convolutional recurrent unit network. This innovative approach aims to develop a highly reliable and accurate spatiotemporal axle temperature forecasting model, thereby increasing locomotive safety and operational efficiency. The modeling structure involves three key steps: (1) the GCN module extracts and aggregates spatiotemporal temperature data and deep feature information from the raw data of different axles; (2) these features are fed into GRU and BiLSTM networks for modeling and forecasting; (3) the ICA algorithm optimizes the fusion weight coefficients to combine the forecasting results from GRU and BiLSTM, achieving superior outcomes. Comparative experiments demonstrate that the proposed model achieves RMSE values of 0.2517 °C, 0.2011 °C, and 0.2079 °C across three temperature series, respectively, indicating superior prediction accuracy and reduced errors compared to benchmark models in all experimental scenarios. The Wilcoxon signed-rank test further confirms the statistical significance of the result improvements with high confidence.

Keywords: spatiotemporal axle temperature forecasting; graph convolutional network; ensemble optimization

1. Introduction

Due to the increasing demand for railway transportation, the reliability and efficiency of railway vehicles have a significant impact on the railway system. The methods applied to ensure the safe operation of railway vehicles have received widespread attention in the development of railway technology [1]. The axle temperature is one of the important indicators that reflect whether the locomotive operation is normal and healthy. Therefore, the prediction of axle temperature is of great value for evaluating the future service status of the bogie and formulating a reasonable operation and maintenance strategy [2]. The axle has a certain temperature fluctuation extent during normal operations. When running under fault conditions, the increase in internal vibration and friction of the axle gearbox would increase the accumulated heat of the faulty axle, resulting in a higher temperature fluctuation range than the normal axles [3]. Therefore, by predicting and tracking the axle temperature of the locomotive bogies, the changing trend of axle temperature in the next period can be anticipated, reserving enough time to address abnormal axle temperatures. This has important engineering value for early warning and adjustment of driving strategy for further decision-making [4].

To achieve accurate and real-time axle temperature monitoring, various axle temperature monitoring and management systems have been developed [5]. Liu developed a monitoring system using onboard switched Ethernet with temperature sensors for temperature fault diagnosis [6]. In the condition monitoring system from Vale et al. [7], different kinds of sensors are utilized in the onboard system to detect faults for early warning. Bing et al. designed a non-destructive embedding measurement method for high-speed trains with axle temperature compensation [8]. Although these systems can obtain relatively accurate real-time measurement data, they still lack the ability to predict future temperature trends. Moreover, the analysis of internal related factors of the system cannot be ignored. Researchers in relevant fields have recently proposed different prediction models for fault diagnosis [9], temperature prediction [10,11], wind speed prediction [12], power prediction [13,14], traffic flow prediction [15,16], air pollutant forecasting [17], etc. Therefore, it is feasible and necessary to establish prediction models for the changing trend of the bogie axle temperature to obtain early warnings and fault diagnosis of the faulty positions on the axle in advance, preventing potential major accidents.

1.1. Related Work

With the advancement of data analysis, time series forecasting methods have gained significant attention, and scholars have designed various models for fault diagnosis in machinery and mechanized equipment [18]. Recently, popular time series models primarily included statistical models and artificial intelligence (AI) models.

Multiple linear regression (MLR) is a statistical method that constructs linear mathematical models to correlate multiple variables and uses sample data for quantitative analysis [19]. Based on the regression analysis, Ma et al. utilized the stepwise regression method to identify important factors in high-speed trains that influence the trend of the collected raw temperature data, by constructing a multivariate forecasting model [1]. The availability of the model has been proven in the forecasting process, but its accuracy remains unsatisfactory. The possible reason may be that the MLR relies on large amounts of statistical data, and the ability of information extraction does not meet application requirements.

With advancements in computing platforms, AI models are widely used due to their superior performance [14]. Hao and Liu applied the backpropagation neural network (BPNN) for high-speed train axle temperature forecasting, showing that the accuracy of BPNN is significantly higher than that of the grey model (GM), GM (1,1) [20]. As parts of AI models, deep learning algorithms are also popular among scholars in axle temperature forecasting. They are able to yield more accurate results than statistical methods by learning from massive data using complex hidden layer structures. Zheng et al. presented the gated recurrent unit (GRU) method for temperature forecasting, showing that GRU had better accuracy than other neural networks [21]. Liu et al. compared the performance of two recurrent neural networks (RNNs), long short-term memory (LSTM) and GRU [22]. Both models achieved accurate predictions of machinery condition data trends and showed a clear advantage in handling non-stationary datasets compared to the autoregressive integrated moving average (ARIMA) model. Yang et al. utilized LSTM for axle temperature forecasting of railway vehicles. The results showed that the framework is feasible and prediction errors are within an acceptable range [23]. Luo et al. also developed an LSTM-based model for train axle temperature prediction using collected data, which showed satisfying results with acceptable error levels [11]. Since the information flow in the LSTM structure moves in one direction only, the optimized bi-directional long short-term memory (BiLSTM) method may be a better solution, more suitable for non-stationary datasets, and able to further improve modeling performance. In the study of Zhang, BiLSTM was employed in a hybrid forecasting framework, proving to have the best robustness and accuracy among benchmark models [24]. The BiLSTM could analyze the non-linearity of raw data to obtain in-depth insights and identify data features.

Based on the above research, deep learning algorithms have proven effective in axle temperature forecasting. However, a single prediction model cannot sufficiently process variable complex nonlinear series data. To improve modeling adaptability and prediction accuracy, feature extraction and ensemble learning methods in a hybrid structure are gradually being proposed.

(1) The non-uniform driving speed of the trains may cause non-stationarity in axle temperature data. To eliminate non-stationarity, feature extraction algorithms are used to conduct the deep information mining of the raw data and extract the most important data features for modeling [25]. Yan et al. applied the stacked autoencoder (SAE) to improve the input of the group method of data handling (GMDH) model. By integrating SAE and support vector machines (SVMs), the hybrid model shows superior experimental results over other benchmark models in axle temperature prediction [26]. The convolutional neural network (CNN) is employed by Fu et al. to extract deep information from raw monitoring data, which effectively improves model ability and enhances the overall forecasting effect of LSTM [27]. Kong et al. developed a CNN-GRU model to achieve condition forecasting of wind turbines. The study proved that CNN can effectively extract the wave information of raw condition data and validated its effectiveness and availability [28]. As the extension to research the spatial connection between the axle measuring points of trains, Man et al. used the graph convolutional network (GCN) [29] for feature extraction, which effectively increased feature extraction ability from the original data and obtained better results with GRU [30].

(2) Ensemble learning with hybrid modeling techniques can collect extracted features and combine different prediction models by adopting a weighted integration principle to achieve improved performance. This can further optimize prediction accuracy and improve modeling adaptability for different datasets [31]. For this purpose, a GA-ANN structure is proposed, in which the forecasting ability of artificial neural network (ANN) was enhanced by genetic algorithm (GA) [32]. Singh used neutrosophic set theory to improve the particle swarm optimization (PSO) algorithm [33]. It was tested with other datasets and this hybrid framework showed significant improvement in accuracy by adding PSO. Pulido et al. also used PSO to optimize the weights of each subnet to rebuild the network [34]. Compared with the traditional network, their ensemble network has better robustness. Another algorithm, the multi-objective grey wolf optimizer (MOGWO), is employed to combine different deep learning prediction models, resulting in better performance than single prediction models for power forecasting [35]. Song et al. presented a novel ensemble model, which uses the grey wolf optimizer (GWO) to combine several neural networks with different structures and achieve better performance than single networks [36]. Li et al. [37] used the imperialist competitive algorithm (ICA) to optimize and integrate multiple extreme learning machines (ELMs) to build forecasting models with excellent performance. The experiments show that the generalization effect of the ICA-ELM method outperformed the classic ELM.

Through the research listed above, the hybrid structures of the abovementioned models could significantly reduce prediction errors. It can also be found that the feature extraction methods can greatly reduce data variability and generate meaningful information. Therefore, the input data for prediction models is further optimized and extracted to strengthen modeling recognition capability. Moreover, ensemble learning algorithms can analyze the correlation in data series and integrate different neural networks to improve model performance with lower error and higher accuracy.

1.2. Novelty of the Study

This study proposes a spatiotemporal forecasting model consisting of the GCN feature extraction module, the GRU prediction module, the BiLSTM prediction module, and the ICA ensemble optimization method for locomotive axle temperature forecasting. The novelty of this study can be summarized as follows:

(1) Different from the traditional single-variable axle temperature forecasting model, this paper proposes a multi-data-driven forecasting model based on spatiotemporal characteristic information. It is able to extract meaningful information from raw locomotive axle temperature data to study temperature fluctuation tendencies for potential failures.

(2) GCN-GRU and GCN-BiLSTM are applied as the main prediction models to construct the spatiotemporal axle temperature prediction model. On one hand, GCN uses the spatial graph network to effectively aggregate data from different nodes and transfer the extracted graph features to GRU and BiLSTM. On the other hand, GRU and BiLSTM networks with special recurrent structures can deeply analyze the extracted feature information by further learning the temporal dependencies, thus building more stable and accurate prediction models.

(3) ICA is applied in the analysis of locomotive axle temperature to optimize fusion weight coefficients of GCN-GRU and GCN-BiLSTM and integrate the forecasting results. ICA has demonstrated its effectiveness with faster convergence speed and better global optimization ability than other methods in modeling optimization. Thus, the internal relationship of the axle temperature series could be deeply analyzed to achieve better forecasting results.

The proposed spatiotemporal axle temperature prediction model can be deployed on computational terminals within locomotive onboard subsystems. By leveraging edge computing, it provides real-time advanced predictions of axle temperatures at key measurement points on the bogie. Additionally, the model can be deployed on high-performance servers within ground subsystems. When axle temperature data from the onboard subsystem is transmitted to the ground subsystem, advanced predictions can be performed, with the results fed back to the locomotive. These predictions of future temperature states enable a comprehensive assessment of temperature trends at various axle temperature measurement points on the bogie. This includes analyzing temperature differences between corresponding axles and between axles and ambient temperatures, facilitating the early detection of potential faults. Consequently, the system can issue fault status warnings, guiding locomotive operation and management. Overall, the locomotive control system equipped with the axle temperature prediction model can issue alerts for potential abnormal temperature rises based on predicted axle temperatures, thereby enabling real-time detection and diagnosis of bogie faults.

The proposed model shows great innovation in the dynamic integration of GCN, GRU, BiLSTM, and ICA. Therefore, it is of great significance to test the applicability and efficiency of the proposed model in axle temperature spatiotemporal forecasting. To reveal its advanced and accurate performance, several models from other researchers are reproduced and compared with the proposed GCN-GRU-BiLSTM-ICA model.

2. Methodology

2.1. Framework of the Proposed Axle Temperature Prediction Model

The overall framework of the proposed spatiotemporal forecasting model is shown in Figure 1, encompassing feature extraction, deep learning prediction models, and ensemble learning algorithms. The details of the framework are listed below:

Part A: The spatial and temporal datasets of raw axle temperature are preprocessed, and the raw data are separated into the training set, validation set, and test set. The end-to-end learning ability of node features and structural information of the GCN is applied to preprocess the spatiotemporal temperature data of the bogie axles. The combinations of GCN-GRU and GCN-BiLSTM are utilized for spatiotemporal axle temperature prediction, and the model parameters are obtained using the training set. GCN is used to extract the spatiotemporal correlation information from the original data and transmit the acquired features to BiLSTM and GRU to obtain the forecasting results.

Part B: ICA is used to integrate GCN-GRU and GCN-BiLSTM by solving and optimizing two weighting coefficients (w_1 and w_2). ICA optimizes the weight coefficients based on the model forecasting results to improve the overall prediction performance.

Part C: The final output of the ensemble model is calculated using the following equation:

$$\hat{o}(t) = w_1\hat{o}_1(t) + w_2\hat{o}_2(t) \tag{1}$$

where w_1 and w_2 are the weight coefficients of GCN-GRU and GCN-BiLSTM, respectively; $\hat{o}_1(t)$ and $\hat{o}_2(t)$ are the predictions of the two models, respectively.

The validation set is utilized to train the ICA, while the test set is used to analyze and evaluate the prediction ability of the proposed GCN-GRU-BiLSTM-ICA model.
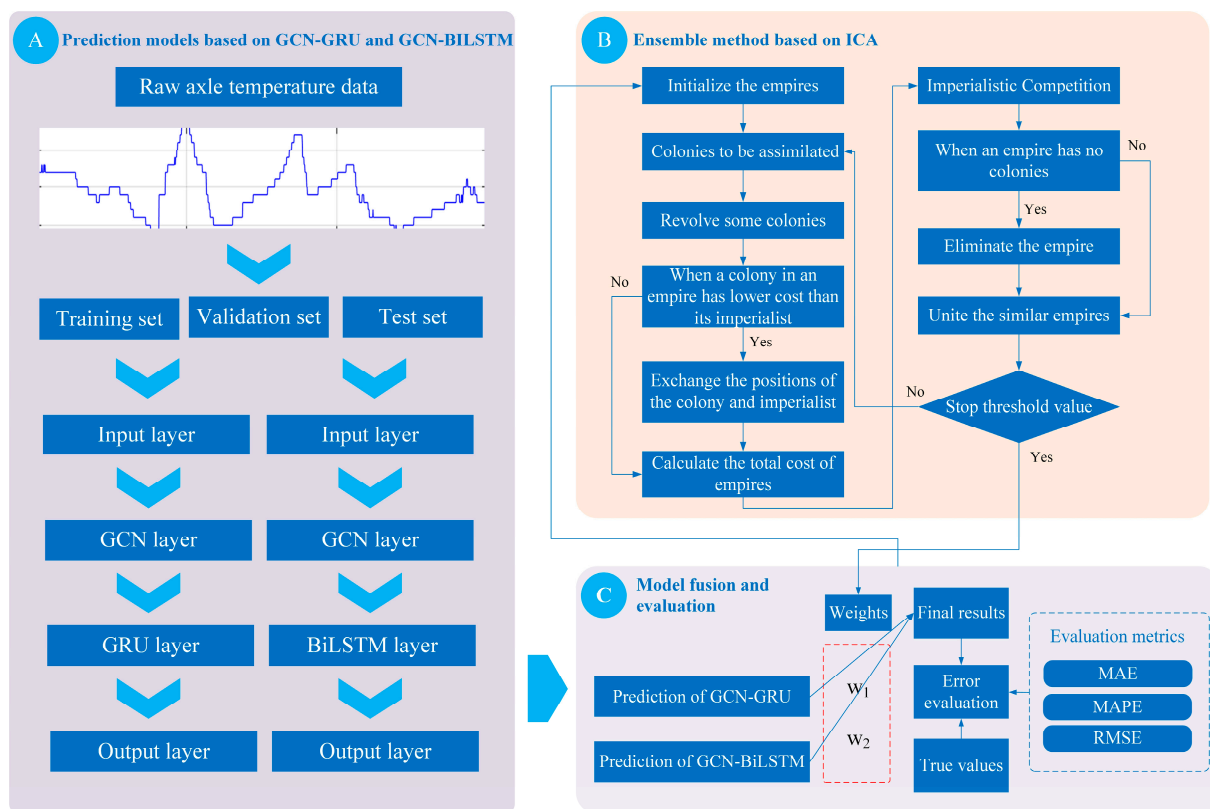

Figure 1. Framework of the proposed GCN-GRU-BiLSTM-ICA model: (**A**) prediction models based on GCN-GRU and GCN-BiLSTM; (**B**) ensemble method based on ICA; (**C**) model fusion and evaluation.

2.2. Graph Convolutional Network

In the analysis of railway vehicles, due to various spatial data such as the complex structure of these vehicles and the relevance of each component, researchers have begun to focus on the application of graph deep learning to extract deep signal features and increase the expressive ability of models on graphs [30]. Graph neural networks can be classified into two categories depending on the type of convolutions: spectral graph neural networks and spatial graph neural networks [38]:

(1) Spectral graph convolution: This method primarily performs convolution operations on topological graphs with the support of graph theory. It transforms data from the spatial domain to the spectral domain by using the graph Fourier transform [39]. The convolution operation is then performed in the spectral domain, and finally, the data are transformed back into the spatial domain using the inverse Fourier transform [40].

(2) Spatial graph convolution: Different from the spectral approach, spatial graph convolution directly operates on the graph structure by defining convolutions on the graph nodes and their neighbors. This method analyzes the correlations of nodes based on their connectivity in the graph, making it more intuitive and often more efficient than spectral methods [41]. Spatial graph convolutions are generally implemented using message-passing techniques, where each node updates its feature representation based on

the information from its neighbors [42]. This method has the advantages of extendibility, high flexibility, and low computational complexity [43]. Hence, spatial graph convolution is applied in this study to describe the locomotive axle temperature data graph network.

To demonstrate graph convolution, the axle temperature graph data can be represented as a graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$, where $\mathcal{V}$ represents the node set, composed of the axle temperature measuring points; $\mathcal{E}$ represents the edge set, containing the connection information between all nodes; $\mathbf{W} \in \mathbb{R}^{N \times N}$ represents the weights of edges. The equations of $\mathbf{W}$ and the degree matrix $\mathbf{D}$ are presented in Equations (2) and (3).

$$\mathbf{W} = \begin{pmatrix} w_{11} & w_{12} & \ldots & w_{1N} \\ w_{21} & w_{22} & \ldots & w_{2N} \\ \ldots & \ldots & \ldots & \ldots \\ w_{N1} & w_{N2} & \ldots & w_{NN} \end{pmatrix} \tag{2}$$

$$\mathbf{D}_{ii} = \sum_{j=1}^{N} w_{ij} \tag{3}$$

where w_{ij} is the weight between node i and node j.

GCN can mine the spatial information of graph structure data. Based on the graph theory, the GCN model can be constructed using the graph Laplacian matrix $\mathbf{L}$, as presented in Equation (4) [44]:

$$\begin{cases} \mathbf{L} = \mathbf{D} - \mathbf{W} \\ \widetilde{\mathbf{L}} = \mathbf{D}^{-\frac{1}{2}}(\mathbf{D} - \mathbf{W})\mathbf{D}^{-\frac{1}{2}} \\ \widetilde{\mathbf{L}} = \mathbf{U}^{\top}\mathbf{\Lambda}\mathbf{U} \end{cases} \tag{4}$$

where $\widetilde{\mathbf{L}}$ represents the symmetric normalization of $\mathbf{L}$; the orthogonal matrix $\mathbf{U}$ and the diagonal matrix $\mathbf{\Lambda} = diag(\lambda_1, \lambda_2, \ldots, \lambda_N)$ represent the eigen decomposition of $\widetilde{\mathbf{L}}$ [45].

Then, the graph convolution can be calculated using Fourier transform and inverse Fourier transform as follows [46]:

$$\mathbf{x} * \mathbf{h} = [\mathbf{U}(\mathbf{U}^{\top}\mathbf{x}) \odot (\mathbf{U}^{\top}\mathbf{h})] \tag{5}$$

where $\mathbf{x}$ is the input graph signal matrix; $\mathbf{h}$ represents the convolutional kernel of graph signals; and $\odot$ represents the Hadamard product.

The GCN can be applied in feature extraction and graph data prediction. An undirected graph is needed for GCN to ensure the decomposition of the graph Laplacian matrix. In the paper, the spatiotemporal axle temperature graph data are represented as an undirected graph due to its structural topology, making it suitable for the application requirements.

2.3. Bi-Directional Long Short-Term Memory

The LSTM, a variety of the RNN, was proposed in 1997 [47]. Compared with other RNNs, the most remarkable feature of LSTM is its gate structure, which can control the information flow [48]. Due to this structure, LSTM is very suitable for the storage and modeling of long historical information and the selective handling of relevant information. The three gates of LSTM are the input gate, output gate, and forget gate, which collaboratively control the flow of information through the LSTM cell [49].

The input and forget gates decide which information should be extracted or discarded, and the output gate analyses the output of the cell. The framework of the LSTM cell is visualized in Figure 2. The transfer and network calculations are shown in the following equations [50]:

$$\mathbf{f}_t = \sigma(\mathbf{W}_f \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f)$$
$$\mathbf{i}_t = \sigma(\mathbf{W}_i \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i)$$
$$\widetilde{\mathbf{r}}_t = \tanh(\mathbf{W}_r \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_r) \qquad (6)$$
$$\mathbf{r}_t = \mathbf{f}_t \odot \mathbf{r}_{t-1} + \mathbf{i}_t \odot \widetilde{\mathbf{r}}_t$$
$$\mathbf{o}_t = \sigma(\mathbf{W}_o \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o)$$
$$\mathbf{h}_t = \mathbf{o}_t \cdot \tanh(\mathbf{r}_t)$$

where $\mathbf{x}_t$ is the input; $\mathbf{f}_t$, $\mathbf{i}_t$, and $\mathbf{o}_t$ are the outputs of the forget gate, input gate, and output gate, respectively; $\mathbf{W}_f$, $\mathbf{W}_i$, $\mathbf{W}_r$, and $\mathbf{W}_o$ are the weight matrices of each gate, respectively; $\mathbf{b}_f$, $\mathbf{b}_i$, $\mathbf{b}_r$, and $\mathbf{b}_o$ are the bias vectors of each gate, respectively; $\mathbf{r}_{t-1}$ and $\mathbf{r}_t$ represent the LSTM cell states; $\mathbf{h}_{t-1}$ and $\mathbf{h}_t$ are the hidden states; σ represents the sigmoid activation function, and tanh represents the hyperbolic tangent activation function; $[\mathbf{h}_{t-1}, \mathbf{x}_t]$ represents the concatenation of $\mathbf{h}_{t-1}$ and $\mathbf{x}_t$.
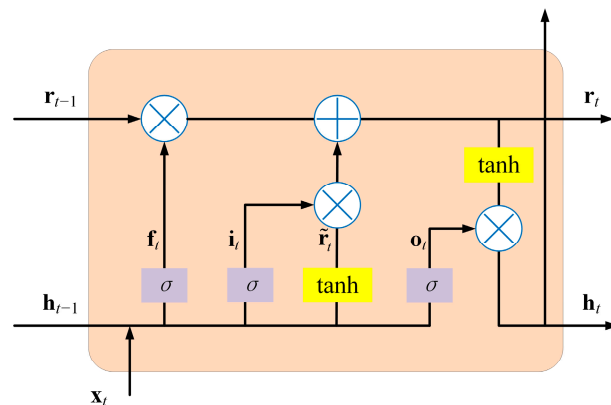

Figure 2. Gate structure and information flow of the LSTM cell.

The BiLSTM consists of a forward LSTM and a backward LSTM, which are usually applied to handle data flow in data processing tasks [51]. LSTM is unable to encode information from back to front whereas BiLSTM can acquire the bi-directional semantic dependencies by connecting two hidden layers with forward and backward data [52]. Generally, BiLSTM can provide the performance of LSTM in prediction and effectively learn deeper correlations. The output layer can obtain bi-directional information on data flow through this framework. The structure of the BiLSTM is demonstrated in Figure 3.
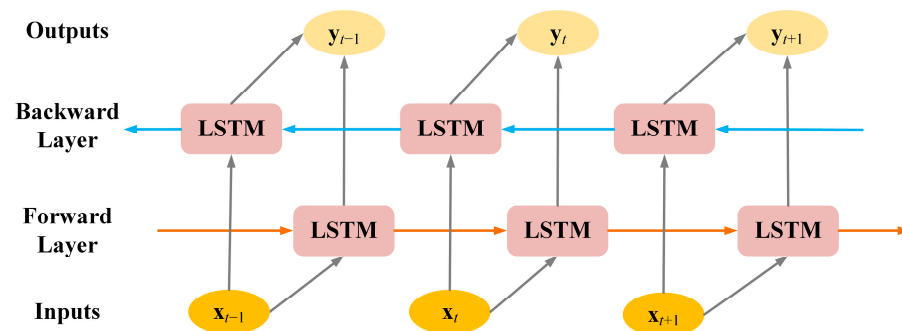

Figure 3. Bi-directional structure of the BiLSTM network.

2.4. Gate Recurrent Unit

The GRU, proposed by Cho et al. in 2014, is a simplified version of LSTM that retains most of the capabilities of LSTM [53]. GRU addresses the long dependency and gradient vanishing problems of the RNN network by storing and processing information flow [54]. The gate structure of the GRU mainly consists of an update gate and a reset gate, resulting in a simpler structure than LSTM and reducing the possibility of overfitting [55].

The update gate determines the amount of historical information to the current status. With a larger value, the update gate can bring more historical information from the previous time step [56]. The reset gate determines the insignificant historical information to ignore. With a smaller value, the reset gate can ignore more historical information [57]. The reset gate functions by analyzing the deep information from the historical data selectively [58]. GRU can be defined as follows [59]:

$$\begin{aligned}
\mathbf{z}_t &= \sigma(\mathbf{W}_\mathrm{z} \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_\mathrm{z}) \\
\mathbf{r}_t &= \sigma(\mathbf{W}_\mathrm{r} \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_\mathrm{r}) \\
\widetilde{\mathbf{h}}_t &= \tanh(\mathbf{W}_\mathrm{h} \cdot [\mathbf{r}_t \odot \mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_\mathrm{h}) \\
\mathbf{h}_t &= (1 - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \widetilde{\mathbf{h}}_t
\end{aligned} \tag{7}$$

where $\mathbf{x}_t$ represents the input; $\mathbf{z}_t$ is the output of the update gate; $\mathbf{r}_t$ is the output of the reset gate; $\widetilde{\mathbf{h}}_t$ is the candidate activation state; $\mathbf{h}_t$ represents the output vector; $\mathbf{W}_\mathrm{z}$, $\mathbf{W}_\mathrm{r}$, and $\mathbf{W}_\mathrm{h}$ are the corresponding weight matrices; $\mathbf{b}_\mathrm{z}$, $\mathbf{b}_\mathrm{z}$, and $\mathbf{b}_\mathrm{h}$ are the bias vectors; σ represents the sigmoid activation function. The structure of the GRU cell is visualized in Figure 4.
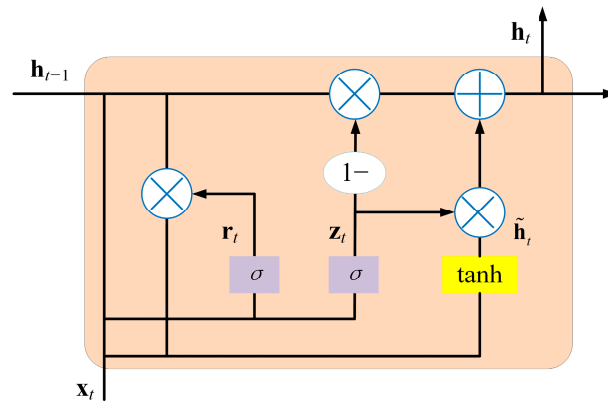

Figure 4. Gate structure and information flow of the GRU cell.

2.5. Imperialist Competitive Algorithm

The ICA, proposed in 2007, is an optimization method formed by simulating the colony assimilation mechanism and the imperial competition system [60]. The ICA draws on the competition, occupation, and annexation of colonies between empires in the political and social colonial times of human history to model the evolution of empires. The ICA is a global optimization algorithm, in which all initialized individuals are divided into two types: imperialists and colonies, based on national power [61].

To handle a multi-dimensional optimization problem, a country can be represented by an array [62]:

$$country = [V_1, V_2, V_3 \ldots, V_N] \in \mathbb{R}^{1 \times N} \tag{8}$$

where V_i represents the i-th variable to be optimized.

The cost of each country is calculated by a cost function f:

$$cost = f(country) = f(V_1, V_2, V_3 \ldots, V_N) \tag{9}$$

At the beginning of the optimization, a total of N_country countries are initialized. The $N_\mathrm{imperialist}$ countries with the least cost and most power are selected as imperialists. The remaining N_colony countries are regarded as colonies.

The normalized imperialist cost C_n for colonization is described as:

$$C_n = c_n - \max_i\{c_i\} \tag{10}$$

where c_n represents the cost of the n-th imperialist and $\max_i\{c_i\}$ is the imperialist with the highest cost.

The normalized power for the imperialist p_n is:

$$p_n = \left| \frac{C_n}{\sum_{i=1}^{N_{\text{imperialist}}} C_i} \right| \tag{11}$$

All imperialists can control some colonies. The number of colonies $N.C_n$ occupied by the *n*-th empire is [63]:

$$N.C_n = \text{round}\left\{ p_n \cdot N_{\text{colony}} \right\} \tag{12}$$

where round denotes a function that rounds numbers. For each imperialist country, it will randomly select some countries from N_{colony} colonies and allocate them to it, forming the initial empires [63].

Assimilation and revolution caused the movement of the colony towards the appropriate empire. If the cost function value of a colony is lower than that of an empire, the colony will switch within the empires [26]. Figure 5 illustrates the movement of the colony by *x* units at an angle of θ:

$$x \sim U(0, \beta \times d) \\ \theta \sim U(-\gamma, \gamma) \tag{13}$$

where *d* represents the distance between imperialist and colony; β $(\beta > 1)$ and γ are random numbers; *U* represents a uniform distribution [64].
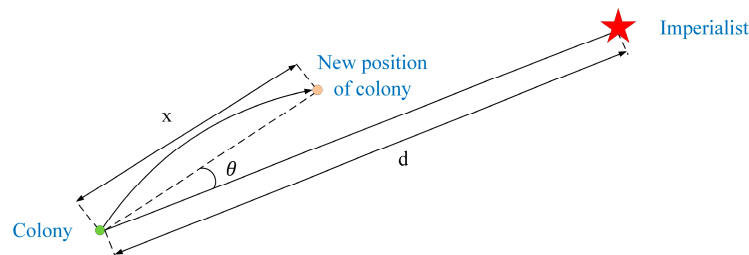

Figure 5. Movement of the colony towards the imperialist.

The total cost $T.C_n$ of the *n*-th empire is:

$$T.C_n = f(imp_n) + \alpha \times \frac{\sum_{i=1}^{N.C_n} f(col_i)}{N.C_n} \tag{14}$$

where the imp_n represents the imperialist country of the *n*-th empire; col_i represents the *i*-th colony of the *n*-th empire; α decides the extent of effect from the colony to the empire.

The competition of empires leads to stronger empires by occupying colonies. An empire will collapse when it possesses no colonies. The ICA stops running when only one empire is left [65].

3. Case Study

3.1. Datasets

To verify the performance and application value of the axle temperature prediction methods for the locomotive status analysis, this paper uses actual axle temperature data collected from a bogie of an electric locomotive for experiments. The data are originally measured and collected by the onboard subsystem of the China locomotive remote monitoring and diagnosis system (CMD).

The bogie consists of three axles, each with two temperature sensors considered. Figure 6a shows the axle and wheel of the bogie of an electric locomotive. Figure 6b shows the spatial structure and topology of the dataset graph, composed of multiple temperature-measuring sensors and their connections. This paper considers six temperature sensors installed on one bogie. The distance between the parallel axles is 1950 mm, and the distance between two measuring points of the same axle is 2120 mm.
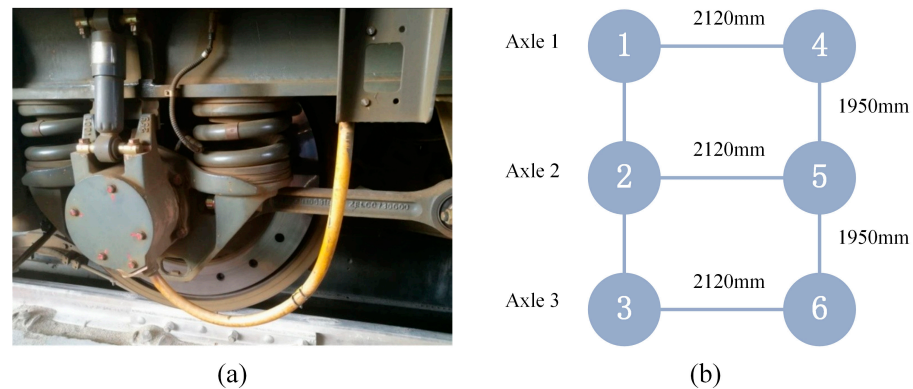

(a)   (b)

Figure 6. Sensor installation and graph structures of the bogie: (**a**) axle and mounted sensors of the bogie; (**b**) spatial structures and graph topology of the temperature measuring points.

In this paper, the sensors are considered the nodes of a graph, and the connections between them are considered the edges of the graph. The time interval between temperature samples is 1 min (thus the sampling rate is 60 samples per hour) and the measurement resolution is 1 °C (i.e., the temperature values are all integers). A total of 1500 samples are collected for each axle. Therefore, a total of 6×1500 samples is used in this paper, and these series are numbered as series #1-#6.

The 1st–900th samples of each series are used as the training set, the 901st–1200th are used as the validation set, and the last 300 samples are used as the test set. In general, the 1st–900th samples are used to train the GCN-GRU and GCN-BiLSTM, the 901st–1200th samples are selected for the optimization of ICA, and the 1201st–1500th samples are used to test the performance of the proposed GCN-GRU-BiLSTM-ICA.

To establish a high-precision spatiotemporal axle temperature prediction model, this paper uses multi-axis graph time series as inputs for the proposed model. The historical axle temperature data collected from six sensors and the graph structure of the sensors are used as model input. The proposed model finally provides the prediction for the point of interest. Measuring points 1, 2, and 3 are selected as three distinct points of interest, which means that the axle temperature series #1, #2, and #3 are separately used for prediction and performance evaluation.

3.2. Evaluation Metrics

The evaluation metrics for regression analysis can comprehensively assess the deviation between the original data and predicted values. In the paper, three classic evaluation metrics—the mean absolute error (MAE), the root mean square error (RMSE), and the mean absolute percentage error (MAPE)—are used to demonstrate modeling accuracy. Moreover, the promoting percentages of evaluation metrics, P_{MAE}, P_{MAPE}, and P_{RMSE}, are also calculated to further analyze the prediction abilities. The equations of the metrics are presented in (15) and (16), respectively.

$$\begin{cases} \text{MAE} = \dfrac{1}{N}\sum\limits_{t=1}^{N}|r_t - \hat{r}_t| \\[2mm] \text{MAPE} = \dfrac{1}{N}\sum\limits_{t=1}^{N}\left|\dfrac{r_t - \hat{r}_t}{r_t}\right| \times 100\% \\[2mm] \text{RMSE} = \sqrt{\dfrac{1}{N}\sum\limits_{t=1}^{N}(r_t - \hat{r}_t)^2} \end{cases} \tag{15}$$

$$\begin{cases} P_{\text{MAE}} = \dfrac{\text{MAE}_1 - \text{MAE}_2}{\text{MAE}_1} \times 100\% \\[2mm] P_{\text{MAPE}} = \dfrac{\text{MAPE}_1 - \text{MAPE}_2}{\text{MAPE}_1} \times 100\% \\[2mm] P_{\text{RMSE}} = \dfrac{\text{RMSE}_1 - \text{RMSE}_2}{\text{RMSE}_1} \times 100\% \end{cases} \tag{16}$$

where r_t is the actual axle temperature at time t; $\hat{r}_t$ is the predicted axle temperature at time t; and N represents the number of samples.

In addition to the aforementioned regression error evaluation metrics, this paper also employs the non-parametric Wilcoxon signed-rank test to verify whether there is a statistically significant difference in the prediction performance between the proposed model and individual models. The null and alternative hypotheses are formulated as follows:

- Null hypothesis (H_0): There is no significant difference in the median prediction errors between model A and model B;
- Alternative hypothesis (H_1): There is a significant difference in the median prediction errors between model A and model B.

3.3. Comparison and Analysis of Different Modules

3.3.1. Comparison and Analysis of Different Prediction Models

To demonstrate the effective prediction performance of the individual GRU and BiLSTM within the proposed hybrid axle temperature forecasting model, several traditional models and classic deep learning models are selected for comparison. These models include RNN, deep belief network (DBN), echo state network (ESN), ELM, multilayer perceptron (MLP), SVM, and ARIMA. The metrics of each model are presented in Table 1 and visualized in Figure 7, from which the following conclusions can be drawn:

(1) The accuracy of the ELM, MLP, SVM, and ARIMA is worse than other models, which might be influenced by the nonstationary and nonlinearity of the raw data series. This demonstrates that deep networks with multiple hidden layers can better analyze the deeply hidden information within graph time series. They outperform traditional shallow neural networks by extracting more data features and achieving superior results.

(2) Compared with other classic deep learning networks, the GRU and BiLSTM show better prediction accuracy. This indicates that these two models can effectively conduct axle temperature forecasting and achieve better results than the other models. The possible reason may be that the GRU improves modeling efficiency through its gate structure, and the bi-directional operation structure of BiLSTM can extract deeper information from both directions of the time series.

(3) It can be observed that the BiLSTM and GRU exhibit satisfying prediction performance compared to other deep learning models. However, they still show different strengths and weaknesses across different series, reflecting the fluctuation among the best prediction accuracy. This proves that a single deep learning prediction model is difficult to adapt to different types of axle temperature series. Therefore, individual deep networks may lead to varying recognition and calculation performances in different time series forecasting. It is necessary to adopt other modules and models to further enhance the prediction ability and robustness of axle temperature forecasting.
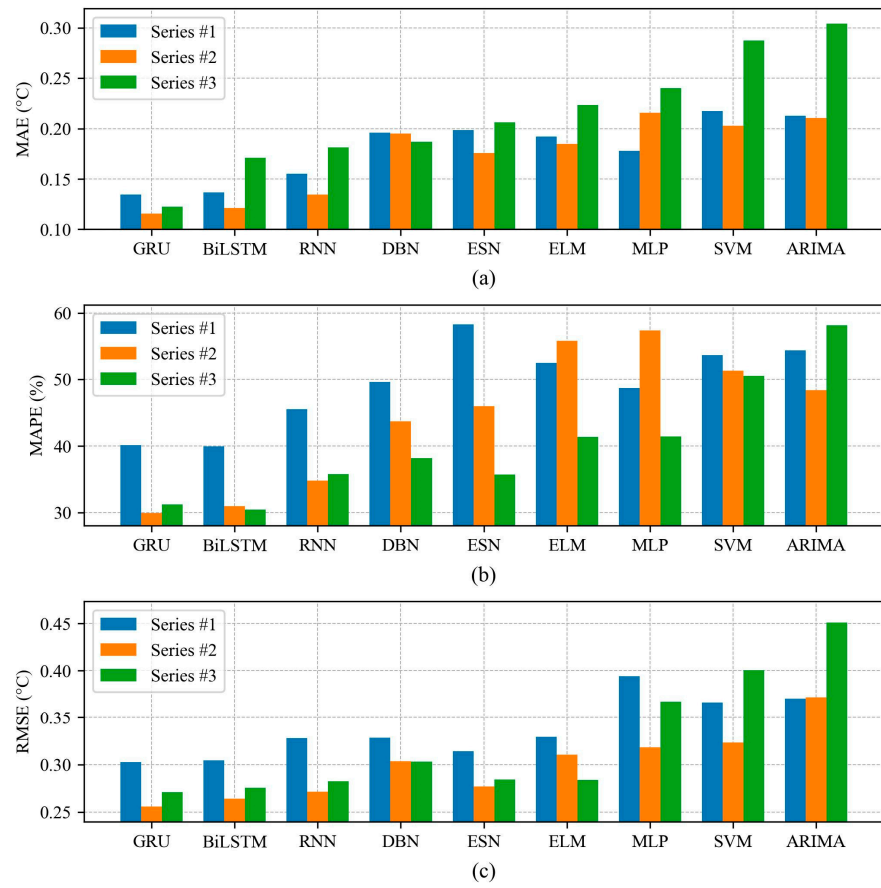
Figure 7. Comparison of error metrics for various individual prediction models on each series: (**a**) MAE metrics; (**b**) MAPE metrics; (**c**) RMSE metrics.

Table 1. Error metrics for various individual prediction models.

Series	Model	MAE (°C)	MAPE (%)	RMSE (°C)
#1	GRU	**0.1345**	0.4013	**0.3029**
	BiLSTM	0.1365	**0.3993**	0.3045
	RNN	0.1550	0.4550	0.3277
	DBN	0.1960	0.4962	0.3282
	ESN	0.1987	0.5830	0.3141
	ELM	0.1920	0.5250	0.3291
	MLP	0.1781	0.4869	0.3941
	SVM	0.2174	0.5367	0.3664
	ARIMA	0.2127	0.5440	0.3705
#2	GRU	**0.1158**	**0.2988**	**0.2561**
	BiLSTM	0.1211	0.3096	0.2645
	RNN	0.1345	0.3479	0.2717
	DBN	0.1950	0.4372	0.3038
	ESN	0.1757	0.4597	0.2770
	ELM	0.1850	0.5582	0.3105
	MLP	0.2156	0.5735	0.3184
	SVM	0.2029	0.5134	0.3231
	ARIMA	0.2107	0.4837	0.3717

Table 1. *Cont.*

Series	Model	MAE (°C)	MAPE (%)	RMSE (°C)
	GRU	**0.1226**	0.3122	**0.2711**
	BiLSTM	0.1710	**0.3043**	0.2759
	RNN	0.1813	0.3577	0.2827
	DBN	0.1868	0.3819	0.3031
#3	ESN	0.2061	0.3571	0.2844
	ELM	0.2235	0.4136	0.2841
	MLP	0.2401	0.4140	0.3672
	SVM	0.2876	0.5057	0.4005
	ARIMA	0.3041	0.5814	0.4506

The values in bold represent the lowest error metrics.

3.3.2. Comparison and Analysis of Different Feature Extraction Modules

To explore the possibility of further optimization of the GRU and BiLSTM, the GCN is applied to extract the graph feature information and is connected with deep learning prediction models to obtain more satisfactory accuracy. The proposed GCN-GRU and GCN-BiLSTM are compared with individual GRU and BiLSTM. Two other feature extraction models, SAE and CNN, are compared with the GCN. All the relevant experimental results are presented in Tables 2 and 3 and visualized in Figure 8, from which the following can be concluded:

(1) In contrast to individual GRU and BiLSTM prediction models, the hybrid model GCN-GRU and GCN-BiLSTM achieve better prediction results with the feature extraction module GCN. The overall results showed that the GCN could deeply analyze and optimize the original graph data features for prediction models, promoting the improvement in the forecasting accuracy of the hybrid model.

(2) In the comparison experiment with the SAE, the GCN based on graph structure information learning presented better performance. This proves that the GCN as a graph feature extraction method, which is a promotion of the CNN based on graphs, is suitable for researching the deep relations between graph nodes to learn the optimal features. The possible cause may be that GCN can improve the applicability of feature extraction by reasonably considering nodes and analyzing the edge weights and inner connections between the nodes.

(3) In the comparison experiment with the CNN, the GCN-based models also obtained better results in the hybrid framework and improved the overall forecasting result of GRU and BiLSTM, which proves that the GCN had excellent node feature learning and analysis ability in spatiotemporal axle temperature modeling. The GCN could achieve end-to-end learning of node features and convolve the topological graph, which can further improve the learning of the weight correlation between nodes, extract the optimal feature based on the data fluctuation trend, and obtain optimal results.

Table 2. Percentage improvement in error metrics of GCN-GRU and GCN-BiLSTM compared to GRU and BiLSTM.

Model	Metric (%)	Series #1	Series #2	Series #3
GCN-GRU	P_{MAE}	43.0483	43.6960	32.2186
vs.	P_{MAPE}	40.8879	46.3520	22.0051
GRU	P_{RMSE}	17.5305	7.1847	19.4024
GCN-BiLSTM	P_{MAE}	43.8828	44.5087	52.3392
vs.	P_{MAPE}	51.2397	46.9315	21.4591
BiLSTM	P_{RMSE}	17.9639	7.1456	17.1801

Table 3. Error metrics for models with different feature extraction modules across three series.

Series	Model	MAE (°C)	MAPE (%)	RMSE (°C)
#1	GCN-GRU	0.0766	0.2011	0.2598
	GCN-BiLSTM	**0.0743**	**0.1947**	**0.2593**
	CNN-GRU	0.1246	0.3325	0.2625
	CNN-BiLSTM	0.1198	0.3203	0.2613
	SAE-GRU	0.1321	0.3540	0.2650
	SAE-BiLSTM	0.1274	0.3417	0.2638
#2	GCN-GRU	**0.0652**	**0.1603**	**0.2377**
	GCN-BiLSTM	0.0672	0.1643	0.2456
	CNN-GRU	0.1070	0.2733	0.2426
	CNN-BiLSTM	0.0724	0.1767	0.2390
	SAE-GRU	0.1145	0.2916	0.2422
	SAE-BiLSTM	0.1141	0.2919	0.2426
#3	GCN-GRU	0.0831	0.2435	0.2685
	GCN-BiLSTM	**0.0815**	**0.2390**	0.2685
	CNN-GRU	0.0994	0.2892	0.2554
	CNN-BiLSTM	0.0906	0.2653	0.2531
	SAE-GRU	0.1032	0.3035	0.2541
	SAE-BiLSTM	0.1015	0.2980	**0.2520**

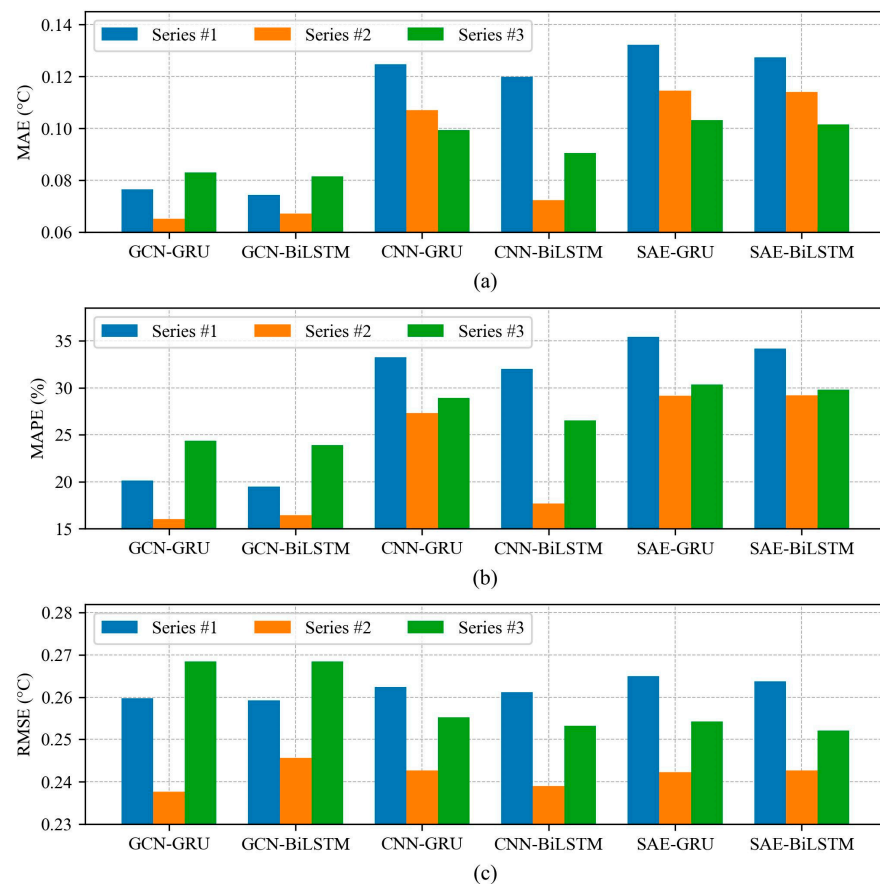The values in bold represent the lowest error metrics.

Figure 8. Comparison of error metrics for models with different feature extraction modules on each series: (**a**) MAE metrics; (**b**) MAPE metrics; (**c**) RMSE metrics.

3.3.3. Comparison and Analysis of Different Ensemble Methods

To verify the effect of ensemble learning on prediction ability and the applicability and effectiveness of the proposed GCN-GRU-BiLSTM-ICA, several ensemble methods are evaluated together with GCN-GRU and GCN-BiLSTM. Moreover, this experiment also verifies the potential of ICA in decision-making by ensemble learning. Several heuristic ensemble methods, such as GWO, GA, and PSO, are used for comparative experiments. The loss convergence of the algorithms is visualized in Figure 9, and the evaluation metrics are presented in Tables 4–6 and visualized in Figure 10, from which the following conclusions can be made:

(1) The prediction accuracies of all ensemble models are superior to the GCN-GRU and GCN-BiLSTM. This proves that ensemble learning algorithms can improve the performance of prediction models. These ensemble learning algorithms can effectively analyze the specialties of the base prediction models and enhance the overall performance of the axle temperature forecasting model.

(2) Among all the ensemble models in this experiment, the ICA algorithm has lower error metrics than other heuristic ensemble methods, proving that the ICA has excellent optimization ability in optimizing the axle temperature prediction models. The possible reason is that the ICA can conduct the optimization with high convergence speed and high convergence accuracy due to its strong global convergence ability. By greatly improving the ability of the exploration analysis and weight coefficients decision, the ensemble model based on ICA can provide optimal predictions.

Table 4. Error metrics for models with different ensemble methods.

Series	Model	MAE (°C)	MAPE (%)	RMSE (°C)
#1	GCN-GRU-BiLSTM-ICA	**0.0725**	**0.1900**	**0.2517**
	GCN-GRU-BiLSTM-PSO	0.0736	0.1929	0.2530
	GCN-GRU-BiLSTM-GA	0.0756	0.1957	0.2590
	GCN-GRU-BiLSTM-GWO	0.0810	0.2155	0.2692
#2	GCN-GRU-BiLSTM-ICA	**0.0633**	**0.1534**	**0.2011**
	GCN-GRU-BiLSTM-PSO	0.0637	0.1557	0.2102
	GCN-GRU-BiLSTM-GA	0.0695	0.1574	0.2159
	GCN-GRU-BiLSTM-GWO	0.0714	0.1663	0.2292
#3	GCN-GRU-BiLSTM-ICA	**0.0789**	**0.2311**	**0.2079**
	GCN-GRU-BiLSTM-PSO	0.0805	0.2360	0.2100
	GCN-GRU-BiLSTM-GA	0.0810	0.2390	0.2154
	GCN-GRU-BiLSTM-GWO	0.0933	0.2426	0.2293

The values in bold represent the lowest error metrics.

Table 5. Percentage improvement in error metrics of GCN-GRU-BiLSTM-ICA compared to GCN-GRU-BiLSTM-PSO, GCN-GRU-BiLSTM-GA, and GCN-GRU-BiLSTM-GWO.

| Model	Metric (%)	Series #1	Series #2	Series #3
GCN-GRU-BiLSTM-ICA vs. GCN-GRU-BiLSTM-PSO	P_{MAE}	1.4946	0.6279	1.9876
	P_{MAPE}	1.5034	1.4772	2.0763
	P_{RMSE}	0.5138	4.3292	1.0000
GCN-GRU-BiLSTM-ICA vs. GCN-GRU-BiLSTM-GA	P_{MAE}	4.1001	8.9209	2.5926
	P_{MAPE}	2.9126	2.5413	3.3054
	P_{RMSE}	2.8185	6.8550	3.4818
GCN-GRU-BiLSTM-ICA vs. GCN-GRU-BiLSTM-GWO	P_{MAE}	10.4938	11.3445	15.4341
	P_{RMSE}	11.8329	7.7571	4.7403
	P_{RMSE}	6.5007	12.2600	9.3328
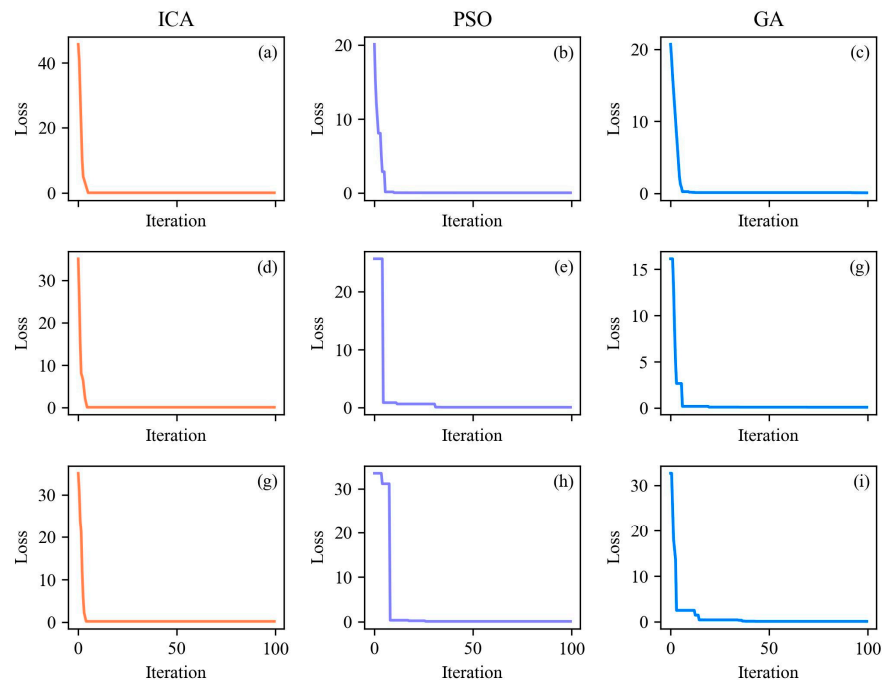
Figure 9. Loss convergence curve of the ICA, PSO, and GA optimization: (**a**) ICA loss on series #1; (**b**) PSO loss on series #1; (**c**) GA loss on series #1; (**d**) ICA loss on series #2; (**e**) PSO loss on series #2; (**f**) GA loss on series #2; (**g**) ICA loss on series #3; (**h**) PSO loss on series #3; (**i**) GA loss on series #3.
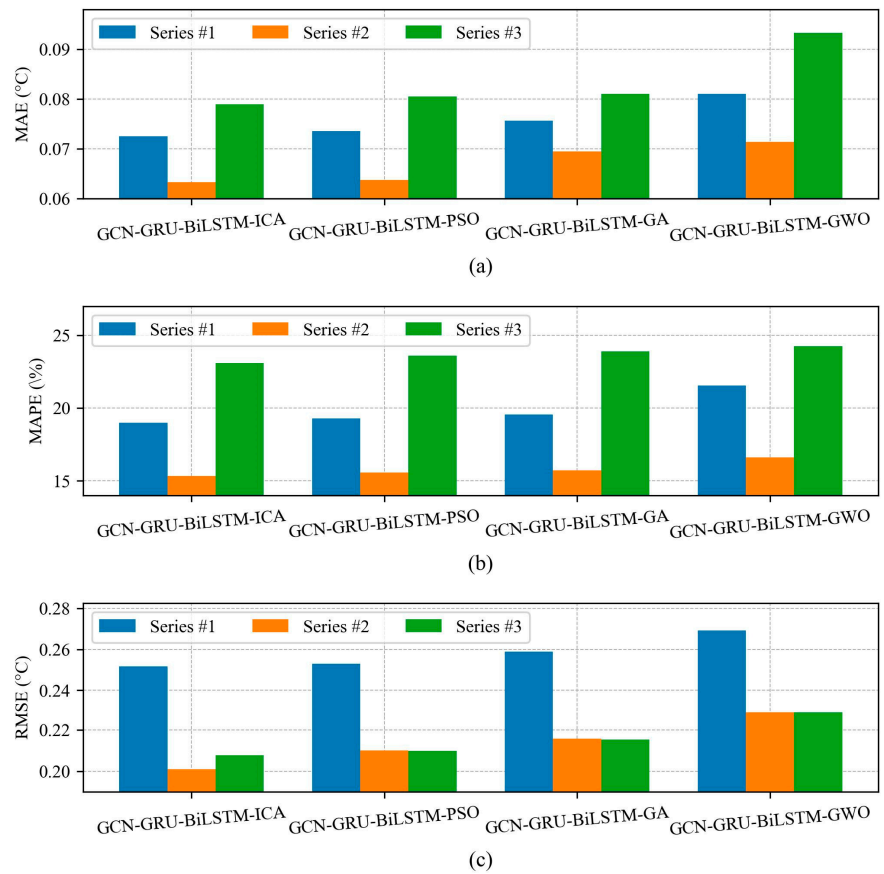

Figure 10. Comparison of error metrics for models with different ensemble methods on each series: (**a**) MAE metrics; (**b**) MAPE metrics; (**c**) RMSE metrics.

Table 6. Percentage improvement in error metrics of GCN-GRU-BiLSTM-ICA compared to GCN-GRU and GCN-BiLSTM.

Model	Metric (%)	Series #1	Series #2	Series #3
GCN-GRU-BiLSTM-ICA	P_{MAE}	5.3168	21.5511	13.5132
vs.	P_{MAPE}	5.0493	7.4407	19.4159
GCN-GRU	P_{RMSE}	5.0397	5.4131	19.5777
GCN-GRU-BiLSTM-ICA	P_{MAE}	5.0322	22.6347	13.4331
vs.	P_{MAPE}	5.3816	6.4984	19.5146
GCN-BiLSTM	P_{RMSE}	5.3727	4.5366	19.6051

3.4. Significance Analysis of the Proposed Model Performance

To evaluate the significance of the performance improvement of the proposed spatiotemporal locomotive axle temperature prediction model over individual baseline models, this section employs the Wilcoxon signed-rank test to analyze the significance of the prediction performance (by using paired prediction errors). Table 7 presents the *p*-values and significance levels of the Wilcoxon tests for comparisons between the proposed model and individual models across different series.

As shown in Table 7, the *p*-values for comparisons between the proposed model and all individual models across different series are far below 0.05. Therefore, we can reject the null hypothesis with a very high level of confidence, concluding that there is a significant difference in predictive performance between the proposed model and each individual baseline model.

Table 7. Wilcoxon *p*-values and significance (α = 0.05) for error comparisons between the proposed model and individual models.

Individual Model	Series #1		Series #2		Series #3	
	p-Value	Significance	*p*-Value	Significance	*p*-Value	Significance
GRU	4.91×10^{-43}	Significant	1.37×10^{-22}	Significant	1.01×10^{-36}	Significant
LSTM	1.02×10^{-42}	Significant	7.45×10^{-38}	Significant	2.26×10^{-39}	Significant
RNN	9.92×10^{-42}	Significant	1.15×10^{-35}	Significant	3.88×10^{-37}	Significant
DBN	1.68×10^{-43}	Significant	7.33×10^{-37}	Significant	7.33×10^{-35}	Significant
ESN	2.96×10^{-43}	Significant	7.12×10^{-40}	Significant	6.60×10^{-36}	Significant
ELM	5.85×10^{-43}	Significant	1.21×10^{-37}	Significant	9.97×10^{-36}	Significant
MLP	3.37×10^{-42}	Significant	2.30×10^{-30}	Significant	7.21×10^{-38}	Significant
SVM	3.22×10^{-43}	Significant	5.17×10^{-39}	Significant	5.86×10^{-32}	Significant
ARIMA	3.87×10^{-44}	Significant	8.06×10^{-40}	Significant	1.41×10^{-41}	Significant

Combining these results with the previously discussed MAE, MAPE, and RMSE evaluation metrics, it is evident that the proposed model significantly outperforms each baseline model in terms of prediction error. The consistent significance across multiple series further reinforces the robustness and generalizability of the performance improvements of the proposed model.

3.5. Comparison and Analysis with Existing Models

The above experiments demonstrate that the proposed GCN-GRU-BiLSTM-ICA axle temperature prediction model achieves relatively high accuracy among all the investigated models. To further verify its effectiveness and innovation, several existing advanced models are used as the benchmark models for comparison, including Mi's model [12], Zhao's model [66], and Liu's model [58]. Additionally, several classical prediction models, such as DBN, SVM, and ARIMA in the time series forecasting field are also included. Figures 11–13 present the forecasting results and residuals of the abovementioned prediction models. Figures 14–16 visualize the MAPE, MAE, and RMSE metrics. From Figures 11–16, the following conclusions can be made:

(1) Among the models mentioned, the prediction performance of advanced models surpasses that of the classical models. This demonstrates that these advanced models are effective in exploring the deep correlation in various features of spatiotemporal axle temperature data and achieving satisfactory results. The possible reason is that these advanced models can efficiently enhance the data aggregation and the overall framework integration degree of deep learning networks through feature extraction or ensemble learning methods.

(2) Compared with the existing models, the proposed GCN-GRU-BiLSTM-ICA model achieves the most precise and satisfactory results in all cases. The proposed GCN utilizes improved topological graphs that establish corresponding relationships between nodes and edges to integrate the input features of six nodes, which increases the quality of the extracted features and the identification power of prediction models. Subsequently, the data after feature extraction are passed into GRU and BiLSTM networks to obtain two forecasting results, respectively. Finally, through dynamic heuristic iteration for weight selection, the ICA effectively combines the results of the two neural networks for the final optimal forecasting results. The proposed model effectively integrates the advantages of each component and demonstrates excellent research potential and application prospects in spatiotemporal axle temperature forecasting.
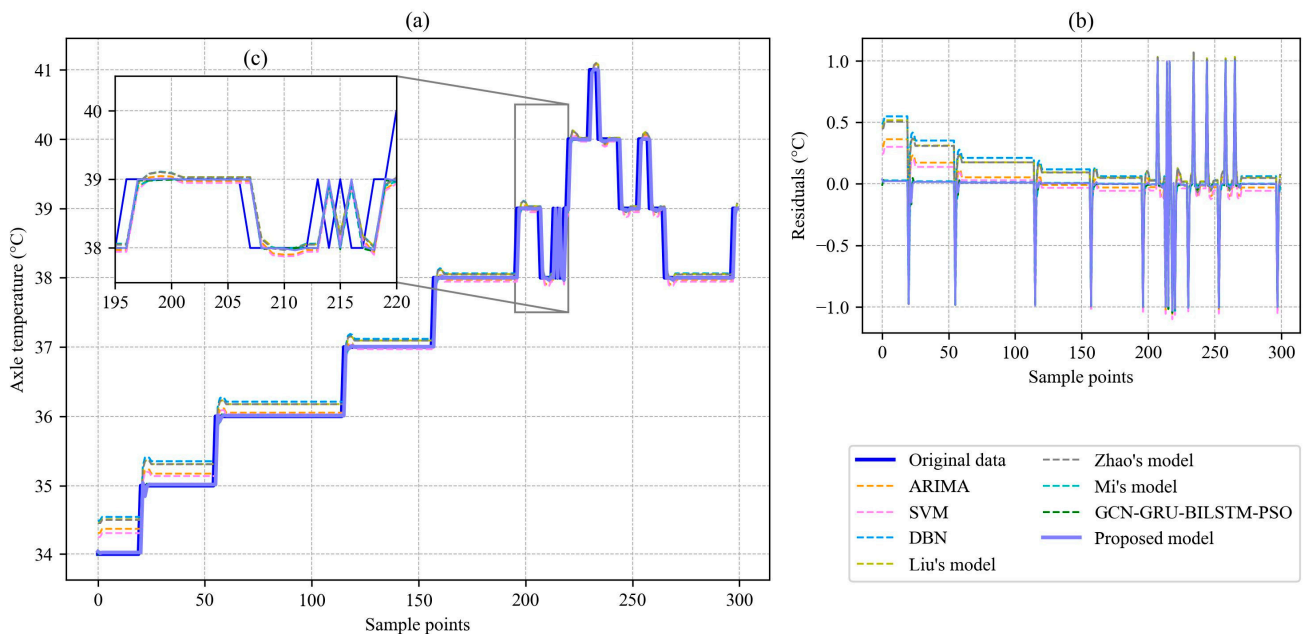

Figure 11. Prediction results and errors of all models for series #1: (**a**) prediction results; (**b**) error distribution; (**c**) local enlargement.
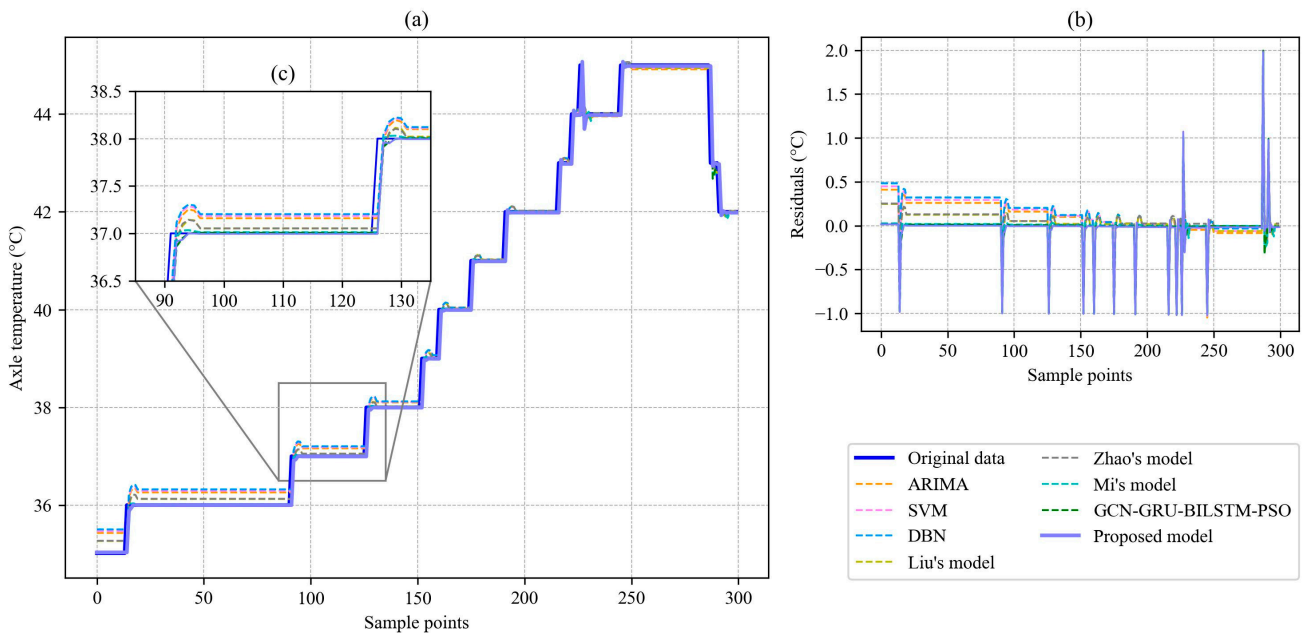
Figure 12. Prediction results and errors of all models for series #2: (**a**) prediction results; (**b**) error distribution; (**c**) local enlargement.
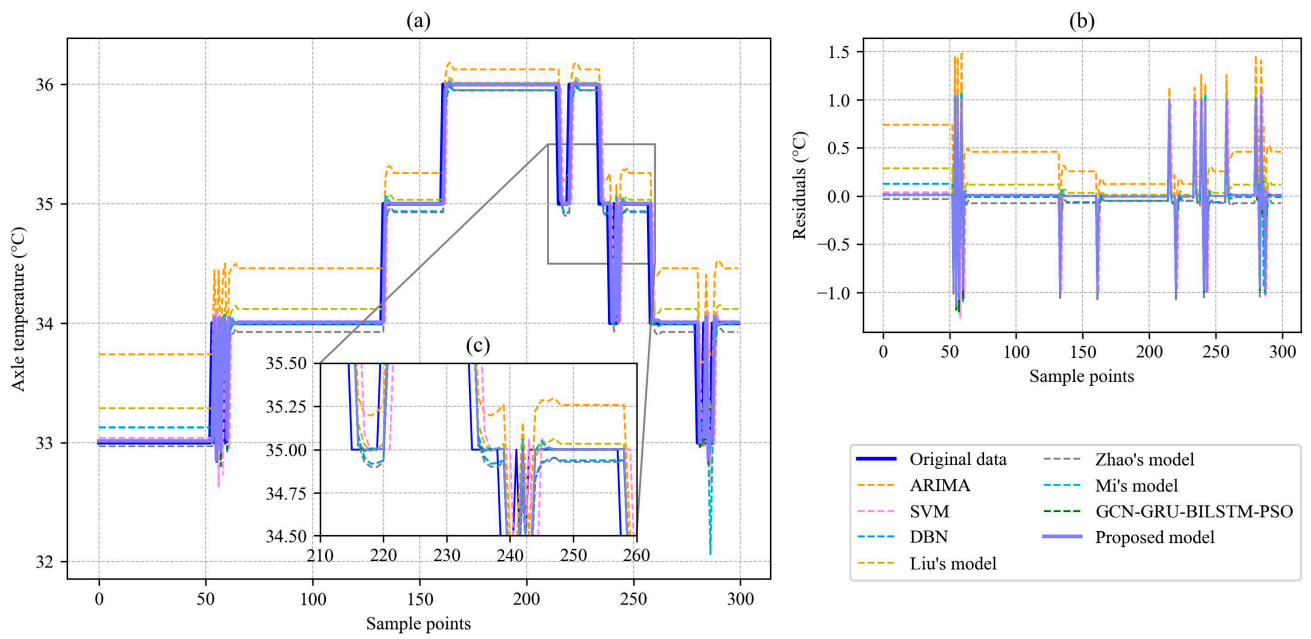

Figure 13. Prediction results and errors of all models for series #3: (**a**) prediction results; (**b**) error distribution; (**c**) local enlargement.
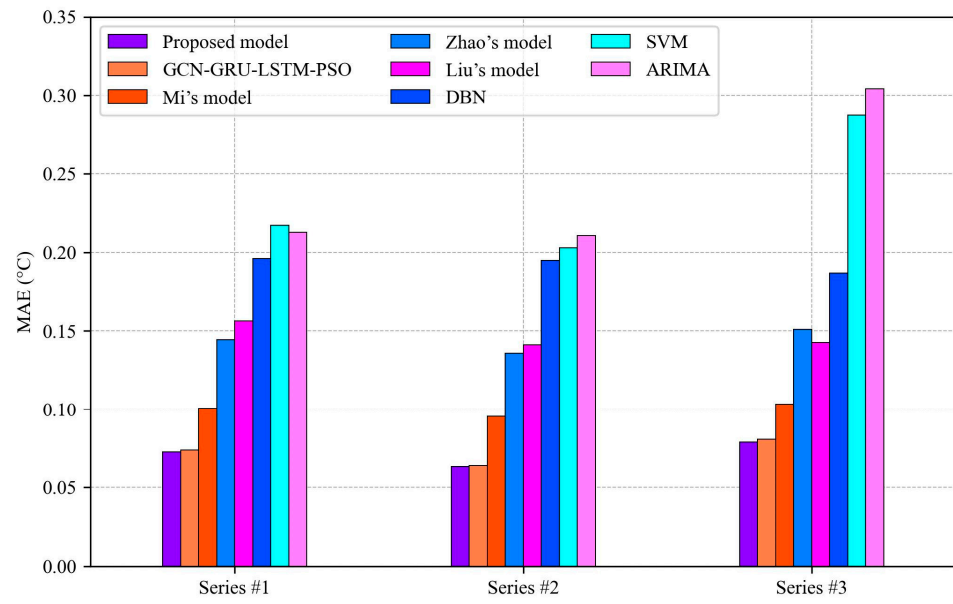
Figure 14. MAE metrics of different prediction models on each series.
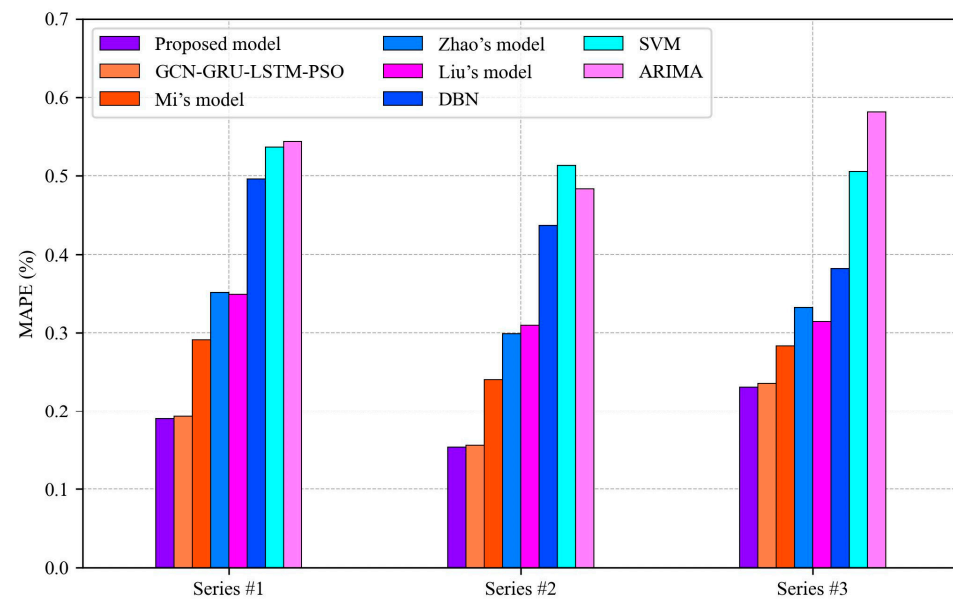

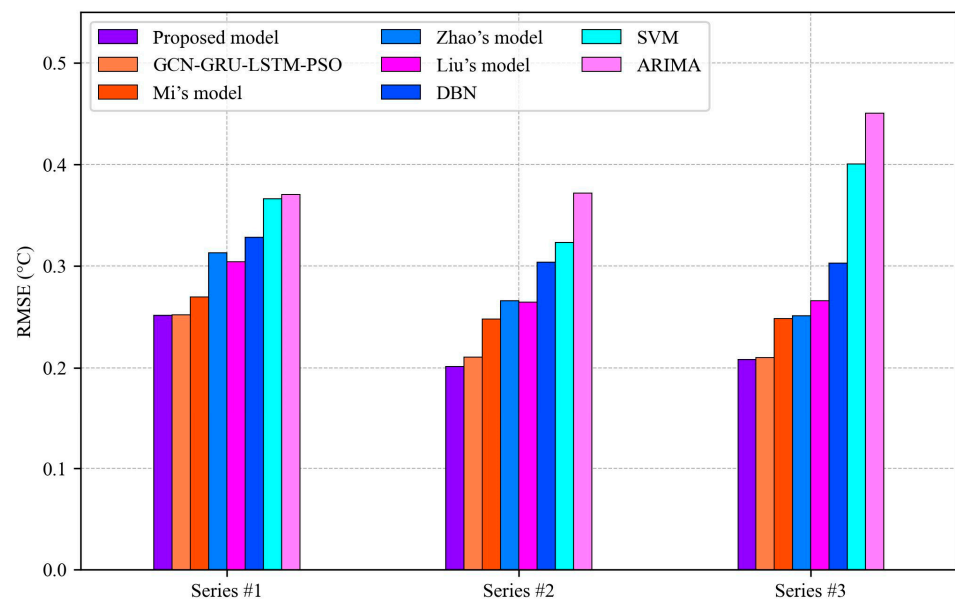**Figure 15.** MAPE metrics of different prediction models on each series.

Figure 16. RMSE metrics of different prediction models on each series.

4. Conclusions

Axle temperature forecasting significantly contributes to real-time failure detection and status management in locomotive control systems. This paper proposes a novel ensemble deep graph prediction model that combines GCN, GRU, BiLSTM, and ICA to construct a multi-data-driven spatiotemporal axle temperature forecasting framework. The key findings of this study are summarized as follows:

(1) The proposed framework integrates data from multiple axle nodes, replacing the typical single-variable time series prediction model. This approach enhances the analysis and identification of axle temperature changes, leading to increased forecasting accuracy.

(2) The GCN module effectively aggregates and extracts spatiotemporal axle temperature data. Unlike traditional feature extraction methods, the graph-based GCN algorithm analyzes spatial correlations among different locomotive axles, extracting spatial-temporal features through a multi-layer graph convolutional neural network. This improves the spatiotemporal modeling capability by providing better input for the prediction models and optimizing the final forecasting results.

(3) The ICA algorithm optimizes the weight coefficients and integrates the forecasting results from GCN-GRU and GCN-BiLSTM, significantly enhancing the applicability and generalization of single deep learning prediction models. The ICA algorithm achieves superior ensemble results compared to other heuristic algorithms by combining effective neural network optimization and decision-making capabilities.

(4) The proposed model framework achieved the best accuracy, with RMSE values of 0.2517 °C, 0.2011 °C, and 0.2079 °C across three temperature series, respectively. The Wilcoxon signed-rank test further confirmed that the performance improvements are statistically significant.

The proposed spatiotemporal axle temperature forecasting model offers valuable insights for locomotive control and intelligent operation management. Future applications could involve embedding the model into onboard subsystems or intelligent ground subsystems to create a comprehensive real-time warning system and lifecycle maintenance scheduling for locomotives.

Nomenclature

AI	Artificial intelligence
ANN	Artificial neural network
ARIMA	Autoregressive integrated moving average
BiLSTM	Bi-directional long short-term memory
BPNN	Backpropagation neural network
CMD	China locomotive remote monitoring and diagnosis system
CNN	Convolutional neural network
DBN	Deep belief network
ELM	Extreme learning machine
ESN	Echo state network
GA	Genetic algorithm
GCN	Graph convolutional network
GM	Grey model
GMDH	Group method of data handling
GRU	Gated recurrent unit
GWO	Grey wolf optimizer
ICA	Imperialist competitive algorithm
LSTM	Long short-term memory
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MLP	Multilayer perceptron
MLR	Multiple linear regression
MOGWO	Multi-objective grey wolf optimizer
PSO	Particle swarm optimization
RMSE	Root mean square error
RNN	Recurrent neural network
SAE	Stacked autoencoder
SVM	Support vector machine

References

1. Ma, W.; Tan, S.; Hei, X.; Zhao, J.; Xie, G. A Prediction Method Based on Stepwise Regression Analysis for Train Axle Temperature. In Proceedings of the 12th International Conference on Computational Intelligence and Security (CIS), Wuxi, China, 16–19 December 2016; pp. 386–390.
2. Wu, S.C.; Liu, Y.X.; Li, C.H.; Kang, G.; Liang, S.L. On the fatigue performance and residual life of intercity railway axles with inside axle boxes. *Eng. Fract. Mech.* **2018**, *197*, 176–191. [CrossRef]
3. Li, C.; Luo, S.; Cole, C.; Spiryagin, M. An overview: Modern techniques for railway vehicle on-board health monitoring systems. *Veh. Syst. Dyn.* **2017**, *55*, 1045–1070. [CrossRef]
4. Milic, S.D.; Sreckovic, M.Z. A Stationary System of Noncontact Temperature Measurement and Hotbox Detecting. *IEEE Trans. Veh. Technol.* **2008**, *57*, 2684–2694. [CrossRef]
5. Singh, P.; Huang, Y.P.; Wu, S.-I. An Intuitionistic Fuzzy Set Approach for Multi-attribute Information Classification and Decision-Making. *Int. J. Fuzzy Syst.* **2020**, *22*, 1506–1520. [CrossRef]
6. Liu, Q. High-speed Train Axle Temperature Monitoring System Based on Switched Ethernet. *Procedia Comput. Sci.* **2017**, *107*, 70–74. [CrossRef]
7. Vale, C.; Bonifácio, C.; Seabra, J.; Calçada, R.; Mazzino, N.; Elisa, M.; Terribile, S.; Anguita, D.; Fumeo, E.; Saborido, C. Novel efficient technologies in Europe for axle bearing condition monitoring—The MAXBE project. *Transp. Res. Procedia* **2016**, *14*, 635–644. [CrossRef]
8. Bing, C.; Shen, H.; Jie, C.; Li, L. Design of CRH axle temperature alarm based on digital potentiometer. In Proceedings of the Chinese Control Conference, Chengdu, China, 27–29 July 2016.

9. Yang, X.; Chen, W.; Li, A.; Yang, C.; Xie, Z.; Dong, H. BA-PNN-based methods for power transformer fault diagnosis. *Adv. Eng. Inform.* **2019**, *39*, 178–185. [CrossRef]

10. Wang, X.; Liu, X.; Bai, Y. Prediction of the temperature of diesel engine oil in railroad locomotives using compressed infor-mation-based data fusion method with attention-enhanced CNN-LSTM. *Appl. Energy* **2024**, *367*, 123357. [CrossRef]

11. Luo, C.; Yang, D.; Huang, J.; Deng, Y.D.; Long, L.; Li, Y.; Li, X.; Dai, Y.; Yang, H. LSTM-Based Temperature Prediction for Hot-Axles of Locomotives. *ITM Web Conf.* **2017**, *12*, 01013. [CrossRef]

12. Mi, X.; Zhao, S. Wind speed prediction based on singular spectrum analysis and neural network structural learning. *Energy Convers. Manag.* **2020**, *216*, 112956. [CrossRef]

13. Gou, H.; Ning, Y. Forecasting Model of Photovoltaic Power Based on KPCA-MCS-DCNN. *Comput. Model. Eng. Sci.* **2021**, *128*, 803–822. [CrossRef]

14. Wang, H.; Li, G.; Wang, G.; Peng, J.; Jiang, H.; Liu, Y. Deep learning based ensemble approach for probabilistic wind power forecasting. *Appl. Energy* **2017**, *188*, 56–70. [CrossRef]

15. Zhang, X.; Zhang, Q. Short-Term Traffic Flow Prediction Based on LSTM-XGBoost Combination Model. *Comput. Model. Eng. Sci.* **2020**, *125*, 95–109. [CrossRef]

16. Dong, S.; Yu, C.; Yan, G.; Zhu, J.; Hu, H. A Novel Ensemble Reinforcement Learning Gated Recursive Network for Traffic Speed Forecasting. In Proceedings of the 2021 Workshop on Algorithm and Big Data, Fuzhou, China, 12–14 March 2021; pp. 55–60.

17. Bai, L.; Liu, Z.; Wang, J. Novel Hybrid Extreme Learning Machine and Multi-Objective Optimization Algorithm for Air Pollution Prediction. *Appl. Math. Model.* **2022**, *106*, 177–198. [CrossRef]

18. Schlechtingen, M.; Santos, I.F. Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. *Mech. Syst. Signal Process.* **2011**, *25*, 1849–1875. [CrossRef]

19. Mashaly, A.F.; Alazba, A.A. MLP and MLR models for instantaneous thermal efficiency prediction of solar still under hyper-arid environment. *Comput. Electron. Agric.* **2016**, *122*, 146–155. [CrossRef]

20. Hao, W.; Liu, F. Axle Temperature Monitoring and Neural Network Prediction Analysis for High-Speed Train under Operation. *Symmetry* **2020**, *12*, 1662. [CrossRef]

21. Zheng, L.; Cao, X.; Chen, F. Main Steam Temperature Prediction Modeling Based on Autoencoder and GRU. *J. Phys. Conf. Ser.* **2020**, *1621*, 012038. [CrossRef]

22. Liu, C.-J.; Liu, H.-T.; Bian, C.; Chen, X.-D.; Yang, S.-H.; Wang, X.-F. Investigation of Time-series Prediction for Turbine Machinery Condition Monitoring. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, *1081*, 012022. [CrossRef]

23. Yang, X.; Dong, H.; Man, J.; Chen, F.; Zhen, L.; Jia, L.; Qin, Y. Research on Temperature Prediction for Axles of Rail Vehicle Based on LSTM. In Proceedings of the 4th International Conference on Electrical and Information Technologies for Rail Transportation (EITRT), Qingdao, China, 25-27 October 2019; Qin, Y., Jia, L., Liu, B., Liu, Z., Diao, L., An, M., Eds.; Springer: Singapore, 2020; pp. 685–696.

24. Zhang, B.; Zhang, H.; Zhao, G.; Lian, J. Constructing a PM2.5 concentration prediction model by combining auto-encoder with Bi-LSTM neural networks. *Environ. Model. Softw.* **2020**, *124*, 104600. [CrossRef]

25. Yuan, H.; Li, J.; Lai, L.L.; Tang, Y. Low-rank matrix regression for image feature extraction and feature selection. *Inf. Sci.* **2020**, *522*, 214–226. [CrossRef]

26. Yan, G.; Yu, C.; Bai, Y. Wind Turbine Bearing Temperature Forecasting Using a New Data-Driven Ensemble Approach. *Machines* **2021**, *9*, 248. [CrossRef]

27. Fu, J.; Chu, J.; Guo, P.; Chen, Z. Condition monitoring of wind turbine gearbox bearing based on deep learning model. *IEEE Access* **2019**, *7*, 57078–57087. [CrossRef]

28. Kong, Z.; Tang, B.; Deng, L.; Liu, W.; Han, Y. Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA data by convolutional neural networks and gated recurrent units. *Renew. Energy* **2020**, *146*, 760–768. [CrossRef]

29. Fu, X.; Pan, Y.; Zhang, L. A Causal-Temporal Graphic Convolutional Network (CT-GCN) Approach for TBM Load Prediction in Tunnel Excavation. *Expert Syst. Appl.* **2024**, *238*, 121977. [CrossRef]

30. Man, J.; Dong, H.; Yang, X.; Meng, Z.; Jia, L.; Qin, Y.; Xin, G. GCG: Graph Convolutional network and gated recurrent unit method for high-speed train axle temperature forecasting. *Mech. Syst. Signal Process.* **2022**, *163*, 108102. [CrossRef]

31. Niu, M.; Hu, Y.; Sun, S.; Liu, Y. A novel hybrid decomposition-ensemble model based on VMD and HGWO for container throughput forecasting. *Appl. Math. Model.* **2018**, *57*, 163–178. [CrossRef]

32. Kouchami-Sardoo, I.; Shirani, H.; Esfandiarpour-Boroujeni, I.; Besalatpour, A.A.; Hajabbasi, M.A. Prediction of soil wind erodibility using a hybrid Genetic algorithm—Artificial neural network method. *CATENA* **2020**, *187*, 104315. [CrossRef]

33. Singh, P. A novel hybrid time series forecasting model based on neutrosophic-PSO approach. *Int. J. Mach. Learn. Cybern.* **2020**, *11*, 1643–1658. [CrossRef]

34. Pulido, M.; Melin, P.; Castillo, O. Particle swarm optimization of ensemble neural networks with fuzzy aggregation for time series prediction of the Mexican Stock Exchange. *Inf. Sci.* **2014**, *280*, 188–204. [CrossRef]

35. Nie, Y.; Jiang, P.; Zhang, H. A novel hybrid model based on combined preprocessing method and advanced optimization algorithm for power load forecasting. *Appl. Soft Comput.* **2020**, *97*, 106809. [CrossRef]

36. Song, J.; Wang, J.; Lu, H. A novel combined model based on advanced optimization algorithm for short-term wind speed forecasting. *Appl. Energy* **2018**, *215*, 643–658. [CrossRef]

37. Li, C.; Zhu, Z.; Yang, H.; Li, R. An innovative hybrid system for wind speed forecasting based on fuzzy preprocessing scheme and multi-objective optimization. *Energy* **2019**, *174*, 1219–1237. [CrossRef]

38. Zhao, G.; Jia, P.; Zhou, A.; Zhang, B. InfGCN: Identifying influential nodes in complex networks with graph convolutional networks. *Neurocomputing* **2020**, *414*, 18–26. [CrossRef]

39. Shang, P.; Liu, X.; Yu, C.; Yan, G.; Xiang, Q.; Mi, X. A new ensemble deep graph reinforcement learning network for spatio-temporal traffic volume forecasting in a freeway network. *Digit. Signal Process.* **2022**, *123*, 103419. [CrossRef]

40. Tang, J.; Liang, J.; Liu, F.; Hao, J.; Wang, Y. Multi-community passenger demand prediction at region level based on spatio-temporal graph convolutional network. *Transp. Res. Part C Emerg. Technol.* **2021**, *124*, 102951. [CrossRef]

41. Yang, L.; Li, W.; Guo, Y.; Gu, J. Graph-CAT: Graph Co-Attention Networks via local and global attribute augmentations. *Future Gener. Comput. Syst.* **2021**, *118*, 170–179. [CrossRef]

42. Taguchi, H.; Liu, X.; Murata, T. Graph convolutional networks for graphs containing missing features. *Future Gener. Comput. Syst.* **2021**, *117*, 155–168. [CrossRef]

43. Peng, H.; Wang, H.; Du, B.; Bhuiyan, M.Z.A.; Ma, H.; Liu, J.; Wang, L.; Yang, Z.; Du, L.; Wang, S. Spatial temporal incidence dynamic graph neural networks for traffic flow forecasting. *Inf. Sci.* **2020**, *521*, 277–290. [CrossRef]

44. Zhang, S.; Tong, H.; Xu, J.; Maciejewski, R. Graph convolutional networks: A comprehensive review. *Comput. Soc. Netw.* **2019**, *6*, 11. [CrossRef]

45. Lee, J.W.; Lee, W.K.; Sohn, S.Y. Graph convolutional network-based credit default prediction utilizing three types of virtual distances among borrowers. *Expert Syst. Appl.* **2021**, *168*, 114411. [CrossRef]

46. Zhou, F.; Yang, Q.; Zhong, T.; Chen, D.; Zhang, N. Variational graph neural networks for road traffic prediction in intelligent transportation systems. *IEEE Trans. Ind. Inform.* **2020**, *17*, 2802–2812. [CrossRef]

47. Hochreiter, S.; Schmidhuber, J. Long Short-term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]

48. Wu, Y.; Yuan, M.; Dong, S.; Lin, L.; Liu, Y. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. *Neurocomputing* **2018**, *275*, 167–179. [CrossRef]

49. Mirza, A.H.; Kerpicci, M.; Kozat, S.S. Efficient online learning with improved LSTM neural networks. *Digit. Signal Process.* **2020**, *102*, 102742. [CrossRef]

50. Yildirim, O.; Baloglu, U.B.; Tan, R.-S.; Ciaccio, E.J.; Acharya, U.R. A new approach for arrhythmia classification using deep coded features and LSTM networks. *Comput. Methods Programs Biomed.* **2019**, *176*, 121–133. [CrossRef]

51. Schuster, M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [CrossRef]

52. Liang, J.; Wang, L.; Wu, J.; Liu, Z.; Yu, G. Elimination of end effects in LMD by Bi-LSTM regression network and applications for rolling element bearings characteristic extraction under different loading conditions. *Digit. Signal Process.* **2020**, *107*, 102881. [CrossRef]

53. Haidong, S.; Junsheng, C.; Hongkai, J.; Yu, Y.; Zhantao, W. Enhanced deep gated recurrent unit and complex wavelet packet energy moment entropy for early fault prognosis of bearing. *Knowl.-Based Syst.* **2020**, *188*, 105022. [CrossRef]

54. Becerra-Rico, J.; Aceves-Fernández, M.A.; Esquivel-Escalante, K.; Pedraza-Ortega, J.C. Airborne particle pollution predictive model using Gated Recurrent Unit (GRU) deep neural networks. *Earth Sci. Inform.* **2020**, *13*, 821–834. [CrossRef]

55. Liu, J.; Wu, C.; Wang, J. Gated recurrent units based neural network for time heterogeneous feedback recommendation. *Inf. Sci.* **2018**, *423*, 50–65. [CrossRef]

56. Adelia, R.; Suyanto, S.; Wisesty, U.N. Indonesian Abstractive Text Summarization Using Bidirectional Gated Recurrent Unit. *Procedia Comput. Sci.* **2019**, *157*, 581–588. [CrossRef]

57. Arias-Vergara, T.; Argüello-Vélez, P.; Vásquez-Correa, J.C.; Nöth, E.; Schuster, M.; Gonzalez-Rátiva, M.C.; Orozco-Arroyave, J.R. Automatic detection of Voice Onset Time in voiceless plosives using gated recurrent units. *Digit. Signal Process.* **2020**, *104*, 102779. [CrossRef]

58. Liu, X.; Qin, M.; He, Y.; Mi, X.; Yu, C. A new multi-data-driven spatiotemporal PM2.5 forecasting model based on an ensemble graph reinforcement learning convolutional network. *Atmos. Pollut. Res.* **2021**, *12*, 101197. [CrossRef]

59. Sun, P.; Boukerche, A.; Tao, Y. SSGRU: A novel hybrid stacked GRU-based traffic volume prediction approach in a road network. *Comput. Commun.* **2020**, *160*, 502–511. [CrossRef]

60. Atashpaz-Gargari, E.; Lucas, C. Imperialist competitive algorithm: An algorithm for optimization inspired by imperialistic competition. In Proceedings of the 2007 IEEE Congress on Evolutionary Computation, Singapore, 25–28 September 2007; pp. 4661–4667.

61. Geetha Devasena, M.; Gopu, G.; Valarmathi, M. Automated and optimized software test suite generation technique for structural testing. *Int. J. Softw. Eng. Knowl. Eng.* **2016**, *26*, 1–13. [CrossRef]

62. Moayedi, H.; Gör, M.; Foong, L.K.; Bahiraei, M. Imperialist competitive algorithm hybridized with multilayer perceptron to predict the load-settlement of square footing on layered soils. *Measurement* **2021**, *172*, 108837. [CrossRef]

63. Khanali, M.; Akram, A.; Behzadi, J.; Mostashari-Rad, F.; Saber, Z.; Chau, K.-W.; Nabavi-Pelesaraei, A. Multi-objective optimization of energy use and environmental emissions for walnut production using imperialist competitive algorithm. *Appl. Energy* **2021**, *284*, 116342. [CrossRef]

64. Gordan, M.; Razak, H.A.; Ismail, Z.; Ghaedi, K. Data mining based damage identification using imperialist competitive algorithm and artificial neural network. *Lat. Am. J. Solids Struct.* **2018**, *15*, e107. [CrossRef]

65. Tien Bui, D.; Shahabi, H.; Shirzadi, A.; Chapi, K.; Hoang, N.-D.; Pham, B.T.; Bui, Q.-T.; Tran, C.-T.; Panahi, M.; Bin Ahmad, B. A novel integrated approach of relevance vector machine optimized by imperialist competitive algorithm for spatial modeling of shallow landslides. *Remote Sens.* **2018**, *10*, 1538. [CrossRef]

66. Zhao, S.; Mi, X. A novel hybrid model for short-term high-speed railway passenger demand forecasting. *IEEE Access* **2019**, *7*, 175681–175692. [CrossRef]