Intelligent Monitoring of Longitudinal Rail Force Using Time Series Forecasting Large Language Models

ABSTRACT

High-speed railways constitute a critical component of modern transportation infrastructure. However, the longitudinal rail force, influenced by cyclic train loads and environmental conditions, poses considerable safety risks if not accurately monitored. Traditional longitudinal rail force detection techniques primarily rely on physical testing and periodic manual inspections, which significantly limit the potential for real-time and continuous monitoring. To address these limitations, this study introduces an innovative approach employing large language models (LLM) to predict longitudinal rail force based on historical monitoring data. The proposed method is validated through extensive long-term field monitoring of longitudinal rail force on high-speed railway lines, thereby confirming its practical applicability. In contrast to conventional Time Series Forecasting Large Language Models (Time-LLM), the proposed method evaluates the prompt-free architecture. When applied to real-world longitudinal rail force data from the Beijing-Shanghai High-Speed Railway, the model achieves an average coefficient of determination (R^2) of 0.932 and a root mean square error (RMSE) of 3.537 kN, outperforming traditional deep learning models. Furthermore, the model exhibits strong robustness under conditions of intermittent data loss. The proposed framework is seamlessly integrated into a localized intelligent system using Langchain-Chatchat, enabling expert-level recommendations based on domain-specific documentation. Overall, this study presents a practical, efficient, and scalable solution for intelligent railway monitoring, offering an advancement toward safer and more intelligent high-speed railway operations.

Keywords: High-speed railway; Infrastructure service safety; Long-term monitoring; Artificial intelligence; Deep learning

1. Introduction

In modern transportation systems, high-speed railways have emerged as a pivotal development direction in rail transport and a core component of contemporary transit networks, owing to their remarkable advantages in speed, comfort, and capacity. The track system, serving as the fundamental technology ensuring train operational safety, comprises various components including rails, fasteners, sleepers, and ballast, and is influenced by the synergistic effects of subgrade, bridges, and tunnels. Among these, the rail, being the critical component in direct contact with trains, is subjected to the dual effects of cyclic train loads and external environmental conditions. This leads to highly unstable longitudinal rail force variations, making it prone to internal damages such as head checks and web cracks, as well as external deformations like rail bending, which may culminate in rail fractures and potentially catastrophic derailments. A case in point is the severe train derailment at Hatfield, UK, in October 2000, which was triggered by rolling contact fatigue cracks in the rails leading to rail breakage [1]. Consequently, the timely and accurate acquisition of the longitudinal rail force state of high-speed railway, enabling real-time and efficient maintenance management, is crucial for ensuring railway operational safety and mitigating accident risks.

In the field of rail detection, non-destructive testing (NDT) techniques are widely adopted for evaluating rail stress and structural conditions. According to their measured physical mechanisms, these techniques can be generally grouped into stress-measuring methods and strain-measuring methods. Stress-measuring methods identify rail stress by detecting intrinsic physical parameter changes, including the Barkhausen method [2], X-ray diffraction [3], ultrasonic guided wave [4,5], and fiber optic sensing [6,7]. Strain-measuring methods estimate rail stress by measuring the deformation or displacement of rail segments, among which the laser observation method is representative [8]. Based on these mature techniques, railway departments in many countries have carried out long-term rail condition monitoring and safety assessment.

High-speed railways impose more stringent requirements on the stability of track structures compared with conventional railways. High-speed railways generally adopt ballastless track structures, which are characterized by high overall rigidity and extremely low tolerance for longitudinal displacement [9]. Once the longitudinal rail force exceeds the design threshold, it will directly lead to irreversible damage such as track slab cracking and debonding between the base slab and the subbase layer, entailing extremely high difficulty and cost for repair. In addition, the continuously welded rails used in high-speed ballastless tracks are longer in length. When temperature changes occur, the longitudinal rail force generated by the thermal expansion and contraction of the tracks is much greater than that of the short rail segments applied in conventional railways [10]. Therefore, the monitoring of high-speed railways demands higher accuracy and timeliness to ensure operational safety.

In recent years, with the continuous advancement of detection technologies, more sophisticated methods have been increasingly applied to achieve comprehensive and refined monitoring of high-speed railway infrastructure. Specifically, researchers have conducted in-depth discussions on sensor deployment schemes for railways under

varying train operating speeds [11], while systematic comparisons have been carried out on monitoring performance across different vibration frequency ranges [12]. Mechanistic analyses have focused on the wheel-rail contact relationship during train operation [13], alongside long-term deformation monitoring and prediction of track slabs [14], as well as identification of abnormal deflections induced by train movement [15]. Throughout the service life of tracks, key technical indicators are subject to rigorous monitoring, encompassing track irregularity identification [16] and track geometry detection [17], among others. Moreover, critical factors influencing track strain such as ambient temperature [18] and ground subsidence [19] have been fully incorporated into analytical frameworks. To ensure the structural integrity and operational safety of high-speed railway infrastructure, auxiliary systems including traction power supply systems [20] are also under continuous surveillance, while rail corrosion detection has been implemented to mitigate potential risks [21]. Additionally, dynamic responses of asphalt support layers [22] and subgrade structures [23] have been systematically analyzed to assess overall structural stability. Notably, the integration of advanced monitoring equipment such as specialized monitoring vehicles [24], inspection robots [25], and smart rail pads [26] has further enhanced the efficiency and accuracy of high-speed railway monitoring practices.

Owing to the superior capability in learning complex patterns and features, deep learning methods have demonstrated exceptional generalization performance when confronted with unknown datasets [27]. In the field of high-speed railway monitoring, various deep learning approaches have been widely adopted and achieved remarkable results, including convolutional neural networks [28], backpropagation neural networks [29], deep reinforcement learning [30], and semi-supervised learning [31]. Furthermore, to enhance monitoring efficiency, lightweight deep learning methods have been applied in high-speed railway monitoring scenarios [32]. However, long-term monitoring data presents critical challenges including long-distance information dependencies and the need for high-precision short-term predictions, where traditional models may suffer from memory deficiencies and insufficient computational capabilities. Therefore, it is imperative to introduce sophisticated deep learning model frameworks to thoroughly capture internal data characteristics and ensure the physical reliability of predictive data.

Building upon this foundation, for railway practitioners it is of paramount importance to establish a rapid, end-to-end intelligent system capable of transforming field measurement data into actionable maintenance. In recent years, Large Language Models (LLM) have emerged as a transformative force, exhibiting remarkable capabilities in language comprehension and processing [33]. LLMs are natural language processing system based on deep learning techniques. Through self-supervised pretraining on massive text corpora, they learn the statistical patterns and semantic representations of language. Their core architecture typically adopts the Transformer framework [34], which employs a self-attention mechanism to capture long-range dependencies and utilizes multi-layer stacking for hierarchical feature extraction. Owing to their outstanding performance, LLMs have been widely applied across various domains, including education [35], healthcare [36], industry [37], and meteorology [38]. Time Series Forecasting Large Language Model (Time-LLM) is a

novel framework that adapts large language models for time series forecasting tasks without modifying the underlying model architecture [39]. It achieves this through a reprogramming approach, where time series data is encoded into text prototypes and enriched by incorporating prompts as prefixes to enhance the contextual input in natural language form. This methodology has demonstrated notable advantages in time series forecasting tasks.

This study pioneers the application of LLMs in the prediction and analysis of high-speed railway longitudinal rail force. Utilizing empirical longitudinal rail force data from the Beijing-Shanghai High-Speed Railway in China, the research focuses on predicting one-day future data based on the preceding ten-day observations. A comprehensive model framework has been developed, incorporating both a Time-LLM and a Local-LLM. The Time-LLM [39] is employed for longitudinal rail force prediction, with the introduction of a prompt-free variant to enhance computational efficiency. Traditional temporal prediction methods, namely Recurrent Neural Network (RNN) [40] and Long Short Term Memory (LSTM) [41], are implemented as benchmarks for comparative performance evaluation. The Local-LLM is specifically designed to analyze prediction outcomes and generate professional maintenance solutions, as illustrated in Fig. 1.

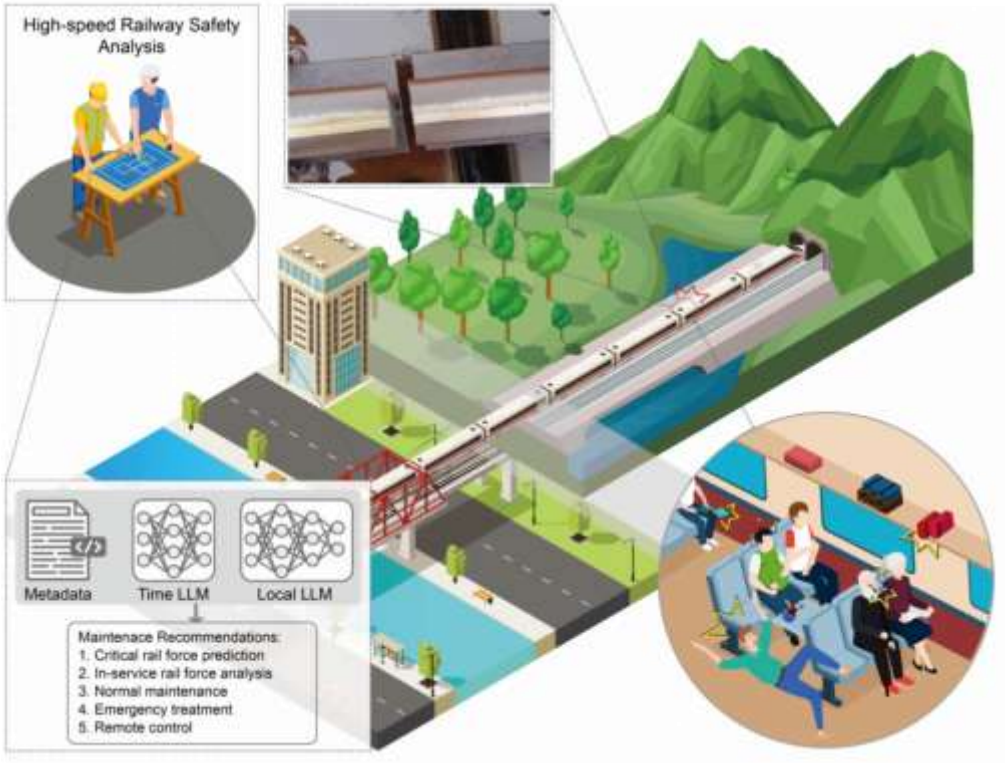


Fig. 1. Large language models for longitudinal rail force monitoring of high-speed railway.

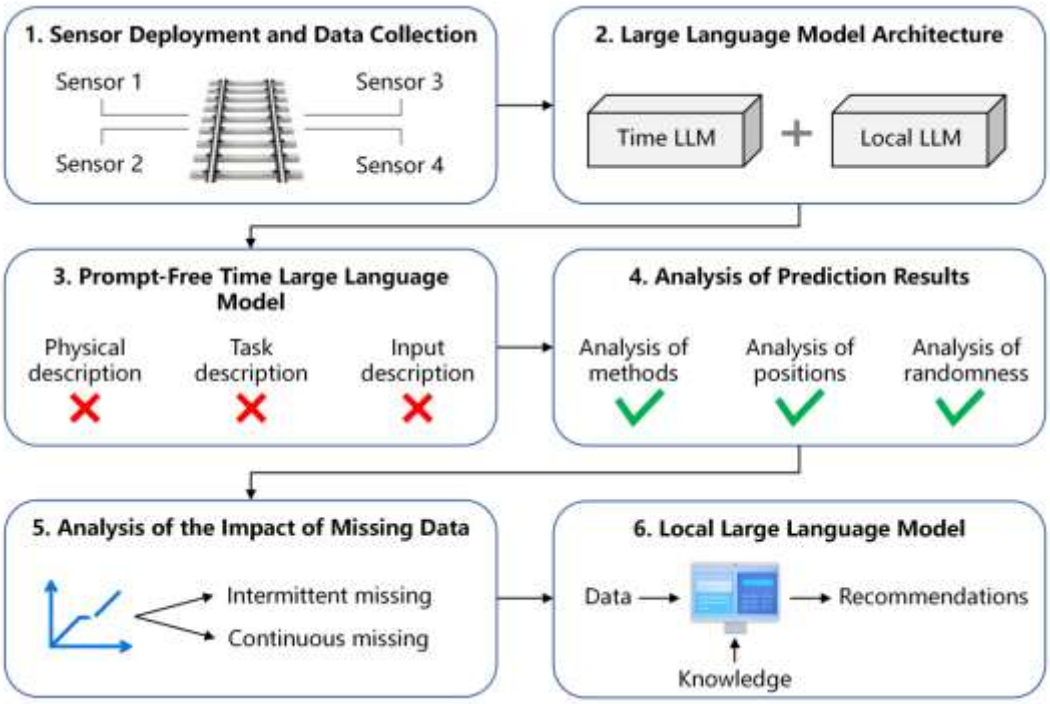


Fig. 2. Flowchart of longitudinal rail force monitoring of high-speed railway.

2. Methods

This study is carried out through the following steps, as shown in Fig. 2.

1. Sensor Deployment and Data Collection: Four sensors are installed at distinct positions within the turnout areas of high-speed railways, with each position exhibiting unique physical characteristics, to collect continuous longitudinal rail force data.

2. Large Language Model Architecture: The LLM includes a Time-LLM [39] and a Local-LLM. The former is used for predicting longitudinal rail force, while the latter generates maintenance recommendations based on the prediction results.

3. Prompt-Free Time Large Language Model: This step discusses the impact of prompt inputs in the Time-LLM on the prediction accuracy and computational efficiency of longitudinal rail force. Three types of prompt inputs, physical description, task description, and input description, are compared both individually and in combination.

4. Analysis of Prediction Results: Based on the prediction results of the prompt-free Time-LLM, this step analyzes its advantages compared with traditional models. Furthermore, the performance and robustness of the prompt-free Time-LLM for sensors at different positions are discussed.

5. Analysis of the Impact of Missing Data: This step evaluates the prediction performance of Time-LLM and prompt-free Time-LLM under different missing data scenarios and discusses the influence of data missing on prediction results.

6. Local Large Language Model: A professional knowledge base for high-speed railways is built within the Local-LLM. Corresponding maintenance recommendations are provided based on the prediction results of the prompt-free Time-LLM and the

professional knowledge base.

2.1 Time Large Language Model

This study employs the Time-LLM framework, a large pre-trained language model specifically developed for time-series data modeling [39]. Its underlying architecture is based on the Transformer framework, widely adopted in deep learning and natural language processing, but incorporates targeted optimizations to accommodate the sequential and temporal nature of time-series data. Through self-supervised pretraining on large-scale temporal datasets, the Time-LLM framework is capable of learning both long-range and short-term dependencies within sequences. During the training phase, the model learns to predict future time-point values based on historical inputs, enabling robust forecasting capabilities. The subsequent content will elaborate on the fundamental principles of Time-LLM and its configuration in this study.

Compared to traditional deep learning approaches, the Time-LLM framework offers significant improvements by integrating the generalization strength of large language models with the specific demands of time-series analysis [39]. Classical architectures such as RNN and LSTM are often challenged by gradient vanishing and struggle with modeling ultra-long dependencies. the Time-LLM framework addresses these challenges by incorporating sparsity-aware attention mechanisms and architectural refinements that enhance its ability to capture long-range dependencies, making it particularly well-suited for modeling complex periodicity, non-stationarity, and structural trends in temporal data.

Furthermore, unlike traditional methods that rely heavily on manual feature engineering and fixed-length inputs, the Time-LLM framework utilizes the representational power of pre-trained models to autonomously extract deep semantic features from raw time-series data [39]. This allows for flexible handling of variable-length sequences, multivariate signals, and incomplete observations, significantly enhancing the model's robustness in processing multi-frequency, heterogeneous datasets. In low-resource or small-sample contexts, where overfitting becomes a critical issue for conventional models, the Time-LLM framework benefits from knowledge transfer through pretraining and prompt-based fine-tuning, enabling it to generalize effectively with minimal labeled data. Moreover, in scenarios involving multimodal data fusion, where traditional time-series models often struggle, the Time-LLM framework’s cross-modal alignment mechanisms facilitate the integration of heterogeneous sources, thereby improving reasoning performance in complex, real-world environments.

At the data processing level, the Time-LLM framework transforms raw time-series inputs into embeddings analogous to word embeddings in natural language processing [39]. Each time point is encoded via an embedding layer that captures fundamental characteristics of the signal. Temporal attributes such as timestamps and periodic features are explicitly encoded to preserve temporal dependencies. The model employs a temporal encoding mechanism similar to positional encoding in Transformers, which includes both absolute and relative time modeling, ensuring sensitivity to temporal order and rhythmicity. This is followed by the application of a multi-head self-attention

mechanism, which enables the model to concurrently attend to diverse temporal patterns across the sequence. This mechanism allows the model to capture fine-grained interactions across different time points, particularly for identifying underlying trends, seasonalities, and anomalies. Each attention layer is succeeded by a feed-forward neural network comprising two fully connected layers, with the second layer incorporating a non-linear activation function to enhance expressive capacity. Finally, a linear projection layer converts the learned hidden representations into predictive values. Mathematically, the predictive process can be formalized as Formula 1.

$$\hat{y}_t = Linear(FFN(MultiHeadAttn(Embed(x_t) + TempEnc(t)))) \quad (1)$$

where $\hat{y}_t$ denotes the predicted value at time t, $Embed(x_t)$ represents the embedding output of the raw time-series input x_t and time t, TempEnc(t) denotes the temporal encoding that models absolute and relative time information, MultiHeadAttn(·) denotes the multi-head self-attention mechanism, FFN(·) represents the feed-forward neural network with non-linear activation, Linear(·) denotes the final linear projection layer.

In consideration of the characteristics of longitudinal rail force monitoring data in high-speed railway systems, this study applies a set of carefully selected hyper-parameters to optimize model performance. The patch length is set to include sixteen time steps per segment, enabling the model to process fixed-size patches suitable for attention-based architectures. A stride of eight time steps is used to define the overlap between adjacent segments, balancing coverage and computational efficiency. The feed-forward network within each Transformer block contains a hidden dimension of 128 units, which enhances non-linear representational capacity while maintaining reasonable complexity. For prediction, only the top candidate from the model's output distribution is selected at each time step to ensure stability and reliability. The hidden layer dimensionality of the underlying language model is configured to 768, allowing for rich temporal representations, while the internal feature representation dimension per time step is set to 128 to control model size. The model employs 128 parallel attention heads, enabling diverse dependency patterns to be captured across the sequence. Both the encoder and decoder operate on seven-dimensional input features, as the same variables are used in both phases. Experimental results from hyper-parameter tuning indicate that this configuration, particularly the use of a large number of attention heads combined with top-score selection, yields optimal prediction performance on the target dataset.

In summary, the Time-LLM framework employs a hierarchical encoder-decoder architecture composed of three core modules: temporal representation learning, cross-modal alignment, and task-specific adaptation [39]. By leveraging the semantic abstraction capabilities of pre-trained language models, it maps raw time-series data into a high-dimensional semantic space and dynamically modulates these representations in accordance with temporal structure. This enables accurate feature extraction and predictive modeling for complex applications such as longitudinal rail force monitoring in high-speed railway systems.

The experiments are performed on a mobile workstation equipped with an 11th

Gen Intel Core i7-11850H CPU and an NVIDIA GeForce RTX 3080 GPU with 16GB of dedicated memory.

2.2 Sensor Deployment and Data Collection

Field monitoring conducted at a switch area of the Beijing-Shanghai High-Speed Railway in China is presented to collect longitudinal rail force data, as shown in Fig. 3. The monitoring site is located in a section of Tianjin where high-speed railway operations are in normal service, ensuring that the study is consistent with actual working conditions. Sensors are fixed on the side of the rail and encapsulated. Given the distinct characteristics of longitudinal rail force variations at different locations, this study examines longitudinal rail force prediction at four strategic positions to enhance the adaptability of the proposed prompt-free Time-LLM [39] model. Sensor 1 is positioned at the straight section near the switch end. Sensor 2 is located at the curved stock rail near the fixed bearing. Sensors 3 and 4 are both situated near the initiation point of the switch rail.

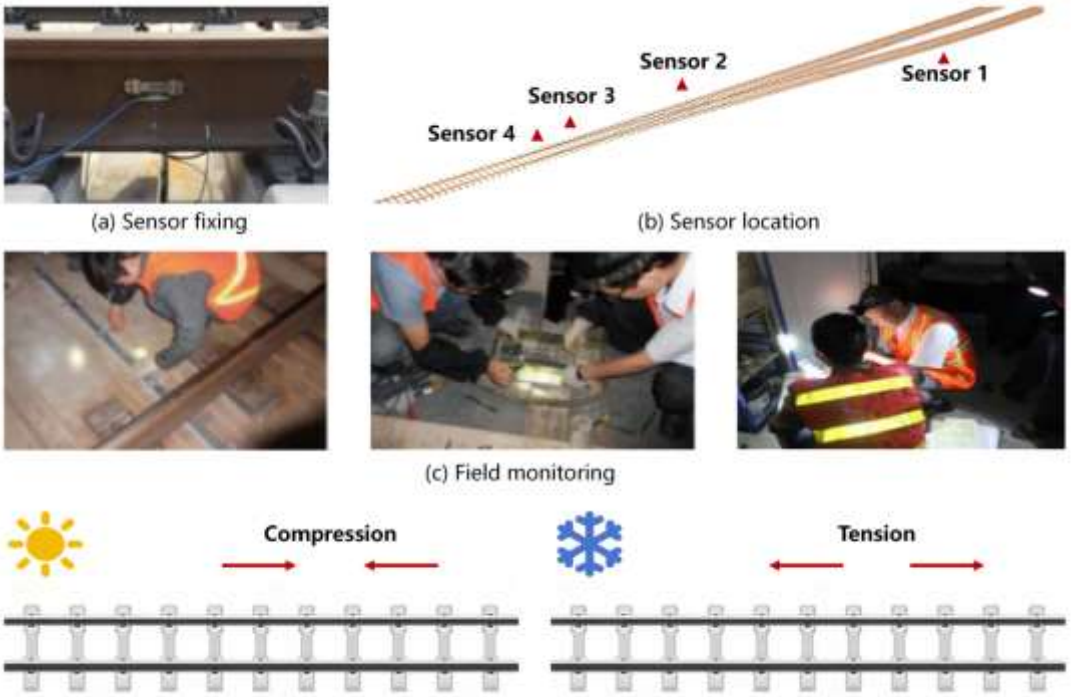


Fig. 3. Sensor deployment and field monitoring.

The sensors adopted in this study are the fiber Bragg grating sensors. The longitudinal rail force alters the grating period, thereby inducing a shift in the fiber Bragg wavelength. Therefore, the variation of the longitudinal rail force can be derived by measuring the wavelength shift. Compared with traditional electrical parameter testing technologies, the fiber Bragg grating sensor boasts such advantages as high measurement accuracy, light weight, compact size and strong electromagnetic interference resistance.

The longitudinal rail force refers to the force acting on the rail along the direction of the rail axis, which is the combined result of multiple factors such as temperature changes, wheel-rail dynamic interactions, and track structure constraints. The

longitudinal rail force is significantly affected by temperature variations. When the temperature is relatively high, the longitudinal rail force manifests as compression; when the temperature is relatively low, it manifests as tension.

The monitoring data analyzed in this study were collected in November 2017. The longitudinal rail force data were collected at a sampling frequency of 15 minutes, yielding 96 data points per day. Aligned with the practical requirements of high-speed railway monitoring projects, the model utilizes a 10-day data set (960 data points) as the training set to predict the subsequent day's longitudinal rail force values (96 data points).

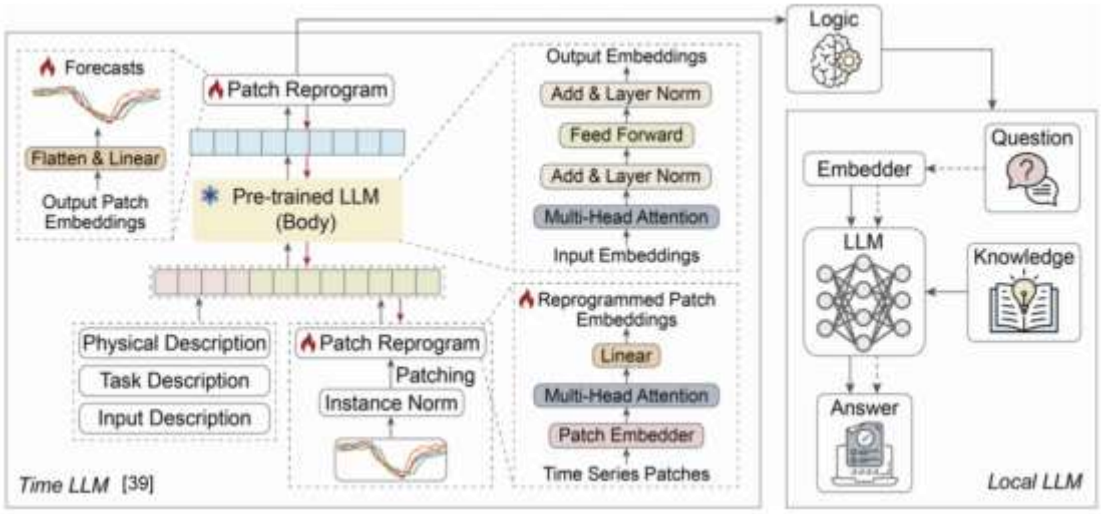


Fig. 4. Structure diagram of the large language model for longitudinal rail force diagnosis of high-speed railway.

2.3 Large Language Model Architecture

The architecture of the LLM proposed in this study for high-speed railway longitudinal rail force diagnosis is depicted in Fig. 4. The system initiates with the Time-LLM module [39], which processes longitudinal rail force data through a sequence of computational steps. The input data first undergoes reprogramming before being transformed into embedding representations. Concurrently, textual prompts are encoded into embeddings via an embedding layer. These two embedding sets are subsequently concatenated and processed by a pre-trained LLM with frozen parameters, where the model weights are initialized using the GPT-2 [42]. The output embeddings from this module are then decoded into predictive results, which are further evaluated through logical reasoning modules to assess potential risks in railway infrastructure.

The risk assessment outputs are subsequently fed into the Local-LLM module. This component can optionally incorporate operational queries related to high-speed railway systems as auxiliary inputs. The Local-LLM converts these multimodal inputs into embedding representations and, through integration with a domain-specific knowledge base, generates comprehensive analytical responses. It is noteworthy that the Local-LLM implementation is based on the LangChain-ChatChat framework [43]. In summary, this research establishes an end-to-end analytical framework for high-speed railway longitudinal rail force prediction and evaluation. The proposed

methodology effectively integrates localized domain expertise with LLM-based predictive modeling, representing an innovative approach that bridges advanced predictive analytics with specialized engineering knowledge in railway infrastructure management. The system demonstrates significant potential for enhancing both the accuracy and interpretability of longitudinal rail force diagnostics through its synergistic combination of temporal pattern recognition and domain-specific knowledge reasoning.

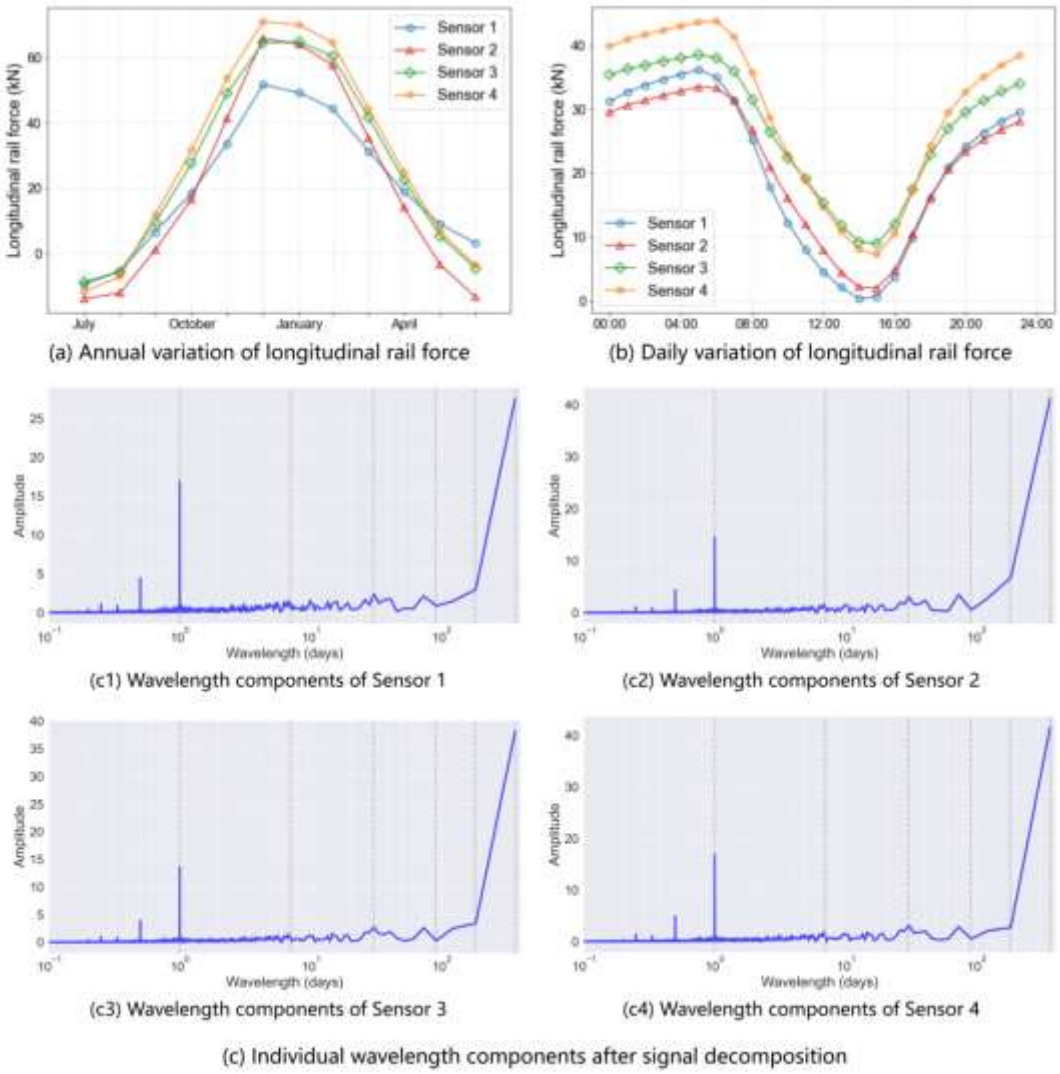
2.4 Prompt-Free Time Large Language Model

The input structure of the Time-LLM comprises of two distinct components: prompt input and time-series input [39]. The prompt input serves to provide supplementary information for the time-series data, including task descriptions and data characteristics. However, the incorporation of prompt input significantly increases computational resource consumption and reduces model training efficiency. In the context of high-speed railway longitudinal rail force prediction, prompt input may not be essential for optimal performance. Consequently, this study implements a prompt-free Time-LLM without prompt input component for predictive analysis and compares its performance with the Time-LLM that includes both input components. This comparative approach enables the evaluation of prompt input's necessity in railway longitudinal rail force prediction tasks.

According to the general architecture of the Time-LLM [39], the prompt input is categorized into three types: physical description, task description, and input description. The physical description is utilized to delineate the physical context, the task description specifies the prediction task information, and the input description characterizes the features of the data. The physical description needs to be tailored according to the specific engineering context, whereas the task description and input description have default settings. In this study, the physical description derived from the deep physical laws analyzed through Fourier transform of the data is considered.

Based on the time-series characteristics of the longitudinal rail force observed by sensors, significant periodic fluctuations are exhibited in the longitudinal rail force on both daily and annual scales. The observation curves of all sensors approximately conform to the variation law of a sine curve, as illustrated in Fig. 5(a) and Fig. 5(b). Tensile longitudinal rail force is defined as a positive value, while compressive longitudinal rail force is defined as a negative value. The underlying cause of these daily and annual periodic variations in the longitudinal rail force lies in the periodic influence of temperature. To quantitatively extract these periodic components, frequency-domain transformation and decomposition are performed on the observed signals. First, the trend term and mean value of the signals are removed to eliminate baseline interference. Subsequently, spectral analysis is employed to identify the significant wavelength components, which represent the dominant periods of the signals and correspond to the angular frequency in the model. As shown in Fig. 5(c), each sensor exhibits six significant wavelength components, among which the most prominent ones correspond to the amplitudes of 1 day and 365 days. This observation is consistent with the characteristic daily and annual periodic variations of the

longitudinal rail force. Based on the aforementioned data characteristics and decomposition results, the observed values Y from the sensors can be described as a linear combination of multiple sinusoidal periodic components, with a regression error term superimposed to account for disturbances. Through this process, the physical description is determined.



(a) Annual variation of longitudinal rail force

(b) Daily variation of longitudinal rail force

(c1) Wavelength components of Sensor 1

(c2) Wavelength components of Sensor 2

(c3) Wavelength components of Sensor 3

(c4) Wavelength components of Sensor 4

(c) Individual wavelength components after signal decomposition

Fig. 5. Analysis of the physical characteristics of longitudinal rail force.

Physical description: "*The sensor observation value \(Y\) can be represented as a regression model composed of specific periodic components. The model is given by \[Y = \varepsilon _ { 0 } + \sum _ { q = 1 } ^ { \mathrm { M } } a _ { q } \cdot \sin \left(\omega _ { q } t + \varphi _ { q } \right)\] Where \(\varepsilon{0}\ )is the regression error, \(a{q}\) are the coefficients of the trigonometric functions, and \(\omega{q}\) represents the angular frequency.*" The corresponding physical formula is shown in Formula 2.

$$Y=\varepsilon_0+\sum_{q=1}^{M} a_q \cdot \sin(\omega_q t+\varphi_q) \tag{2}$$

where Y is the observation value, ε_0 denotes the regression error term, a_q represents the amplitude coefficients of the trigonometric components, ω_q corresponds to the angular frequency for the q-th component, φ_q indicates the phase shift of each sinusoidal function, M is the total number of harmonic components in the model.

For the task description, the default setting: "*forecast the next {str(self.h)} steps given the previous {str(self.input_size)} steps information.*" Regarding the input description, the default configuration: "*min value {min_values_str},max value {max_values_str}, median value {median_values_str}, the trend of input is {'upward' if trends[b] > 0 else 'downward'},top 5 lags are : {lags_values_str}*".

3. Results and discussion

3.1 Comparative Experiment Results

To validate the effectiveness of the prompt, a comparative experiment was conducted, as detailed in Table 1. In this experimental design, Test_0 serves as a blank control to evaluate the performance of the prompt-free Time-LLM [39]. Test_1 is designed to assess the effect of utilizing only the physical description. Test_2 examines the outcome with the task description alone. Test_3 investigates the impact of employing solely the input description. Test_4 is implemented to evaluate the combined effect of simultaneously incorporating physical description, task description, and input description. In the above experiments, the prompts are all structured inputs.

The comparative experimental results demonstrate that the predictive performance of the Time-LLM [39] with prompt input does not surpass that of prompt-free Time-LLM. This finding indicates that prompt input is unnecessary for high-speed railway longitudinal rail force prediction tasks, as the raw data inherently contains sufficient information that can be effectively learned without additional textual input. Notably, Test_4 exhibits significantly inferior performance compared to the other four experimental groups, suggesting that excessive prompts can degrade Time-LLM performance. The reason is that an abundance of prompts introduces redundant information, consequently reducing its capability to extract essential information from the data and leading to diminished predictive accuracy.

The impact of the weights adopted by the prompt-free Time-LLM [39] on longitudinal rail force prediction is analyzed. All the results presented previously are based on GPT-2 weights [42], whereas Llama-2 weights [44] are employed herein for comparative analysis, as shown in Table 2. The comparison results demonstrate that the average R^2 achieved with Llama-2 weights is 0.927, which is slightly lower than the value of 0.932 obtained with GPT-2 weights. However, the gap between the two scores is extremely narrow, accounting for merely 0.5% of the GPT-2-based result. This finding indicates that different weights exert an insignificant influence on the

performance of the prompt-free Time-LLM in longitudinal rail force prediction. Therefore, the prompt-free Time-LLM based on GPT-2 weights is uniformly adopted for all subsequent analyses in this paper.

Table. 1. Comparative experimental settings and results.

Experiments	Physical description	Task description	Input description	Average R^2
Test_0				0.932
Test_1	√			0.931
Test_2		√		0.922
Test_3			√	0.931
Test_4	√	√	√	0.903

Table. 2. The predicted R^2 values of prompt-free Time-LLM based on different weights.

Weights	Sensors				
	1	2	3	4	Average
GPT-2	0.957	0.927	0.910	0.935	0.932
Llama-2	0.936	0.928	0.915	0.929	0.927

3.2 Analysis of Prediction Results

The predictive performance of the lightweight Time-LLM [39] in estimating longitudinal rail force was evaluated through two key metrics: the coefficient of determination (R^2) and root mean square error (RMSE), with detailed comparisons made against conventional LSTM and RNN models in Table 3. Demonstrating better modeling capability, the lightweight Time-LLM achieved an average R^2 value of 0.932 across four different rail locations, outperforming traditional LSTM and RNN models which attained values of 0.781 and 0.537 respectively. Furthermore, the model exhibited enhanced predictive accuracy with an RMSE as low as 3.537 kN, which is lower than the 6.446 and 9.528 kN observed in LSTM and RNN models. These results establish the lightweight Time-LLM as a particularly effective solution for high-speed railway longitudinal rail force prediction, offering substantial improvements over existing approaches in terms of prediction precision, thereby presenting a valuable advancement for railway infrastructure monitoring and maintenance strategies.

The analysis of longitudinal rail force prediction accuracy across different rail locations reveals notable variations that exceed those attributable to computational randomness. These discrepancies primarily stem from distinct longitudinal rail force variation characteristics influenced by location-specific physical factors. Fig. 6(a) shows the prediction performance of longitudinal rail force at different sensors over the next day. Notably, the prompt-free Time-LLM [39] demonstrates smaller variations in

prediction accuracy across different locations compared to both LSTM and RNN. This reduced sensitivity to positional factors and associated physical influences suggests that the prompt-free Time-LLM possesses enhanced applicability and reliability in diverse operational scenarios.

Table. 3. R^2 and RMSE of longitudinal rail force prediction.

	Models	Sensors				
		1	2	3	4	Average
R^2	Prompt-free Time-LLM	0.957	0.927	0.910	0.935	0.932
	LSTM	0.780	0.708	0.805	0.832	0.781
	RNN	0.358	0.563	0.614	0.614	0.537
RMSE (kN)	Prompt-free Time-LLM	3.085	3.586	3.731	3.746	3.537
	LSTM	7.094	7.223	5.489	5.978	6.446
	RNN	12.067	8.919	7.768	9.360	9.528

The comparative analysis of longitudinal rail force prediction results across different random seeds reveals that the prompt-free Time-LLM [39] exhibits R^2 standard deviations ranging from 0.0236 to 0.0282. In contrast, the LSTM demonstrates R^2 standard deviations between 0.062 and 0.082, while RNN shows even greater variability with R^2 standard deviations ranging from 0.049 to 0.268. These values significantly exceed those of the prompt-free Time-LLM, indicating superior prediction stability of the proposed model. An intuitive comparison is shown in Fig. 6(b). Furthermore, the prompt-free Time-LLM maintains RMSE standard deviations between 0.553 kN and 0.855 kN, consistently outperforming both LSTM and RNN in terms of result stability. These findings underscore the robust and stable predictive performance of the prompt-free Time-LLM.

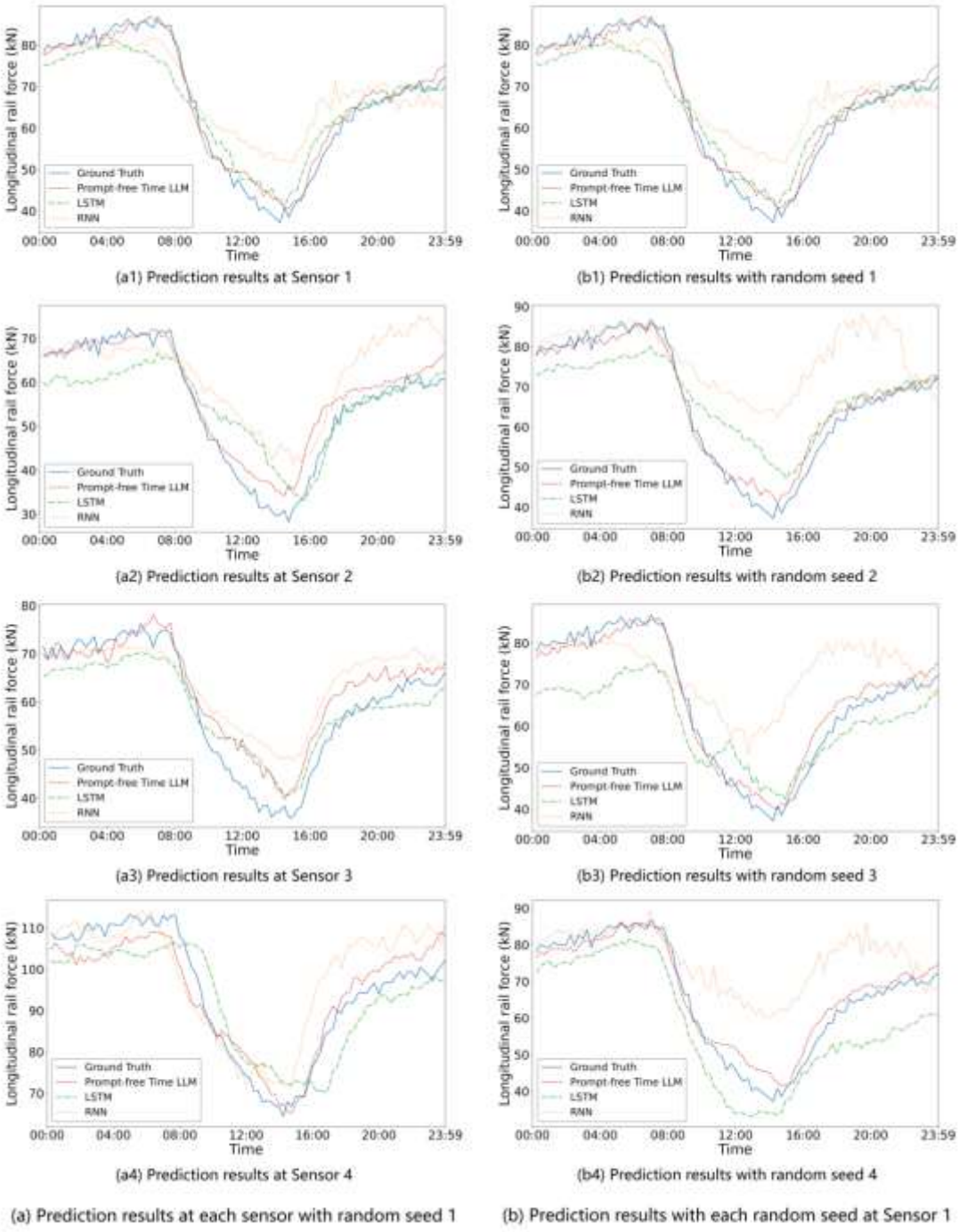


Fig. 6. Comparison of prediction results between prompt-free Time-LLM, LSTM, and RNN.

3.3 Analysis of the Impact of Missing Data

In practical high-speed railway monitoring engineering, data loss frequently occurs due to monitoring equipment anomalies and other factors [45]. To investigate the predictive performance of Time-LLM [39] under different data loss scenarios, this study designed two distinct data loss conditions: (1) 50% intermittent missing data, and (2) 50% continuous missing data. The predictive results obtained using both Time-LLM and prompt-free Time-LLM under these conditions are shown in Table 5.

According to the results presented in Table 5, 50% intermittent missing data leads to a minor decrease in prediction accuracy. This phenomenon can be attributed to the preservation of data trends despite the missing values, allowing both Time-LLM and prompt-free Time-LLM to extract essential variation patterns and maintain predictive capability. Conversely, 50% continuous missing data significantly compromises data characteristics, resulting in substantial deterioration of prediction accuracy. These findings underpin the critical importance of comprehensive data collection practices. Furthermore, the results demonstrate that prompt input cannot compensate for the performance degradation caused by compromised data quality when fundamental data characteristics are disrupted. This observation further validates the universal applicability of prompt-free Time-LLM in various data scenarios.

Table. 5. Longitudinal rail force prediction R^2 under different data missing situations.

Sensor	Model	No missing data	50% intermittent missing data	50% continuous missing data
1	Time-LLM	0.917	0.905	<0
	Prompt-free Time-LLM	0.957	0.948	<0
2	Time-LLM	0.883	0.862	<0
	Prompt-free Time-LLM	0.927	0.893	<0
3	Time-LLM	0.897	0.881	<0
	Prompt-free Time-LLM	0.910	0.898	<0
4	Time-LLM	0.903	0.874	<0
	Prompt-free Time-LLM	0.935	0.902	<0

3.4 Local Large Language Model

The localization was implemented using Langchain-Chatchat [43]. A data input interface was integrated into Langchain-Chatchat, enabling the import of high-speed railway longitudinal rail force monitoring data in CSV format. Prior to this process, a substantial volume of high-speed railway literature had been stored in the knowledge base. Subsequently, Langchain-Chatchat automatically executes the prompt-free Time-LLM [39]. The prediction results are used for retrieval-augmented generation. Specifically, the prediction results are converted into vectors to retrieve relevant documents from the high-speed railway professional knowledge base. These retrieval results are then combined with the prediction results to form an enhanced prompt. Finally, the LLM generates recommendations for high-speed railway operations based

on the enhanced prompt, as shown in Fig. 7.

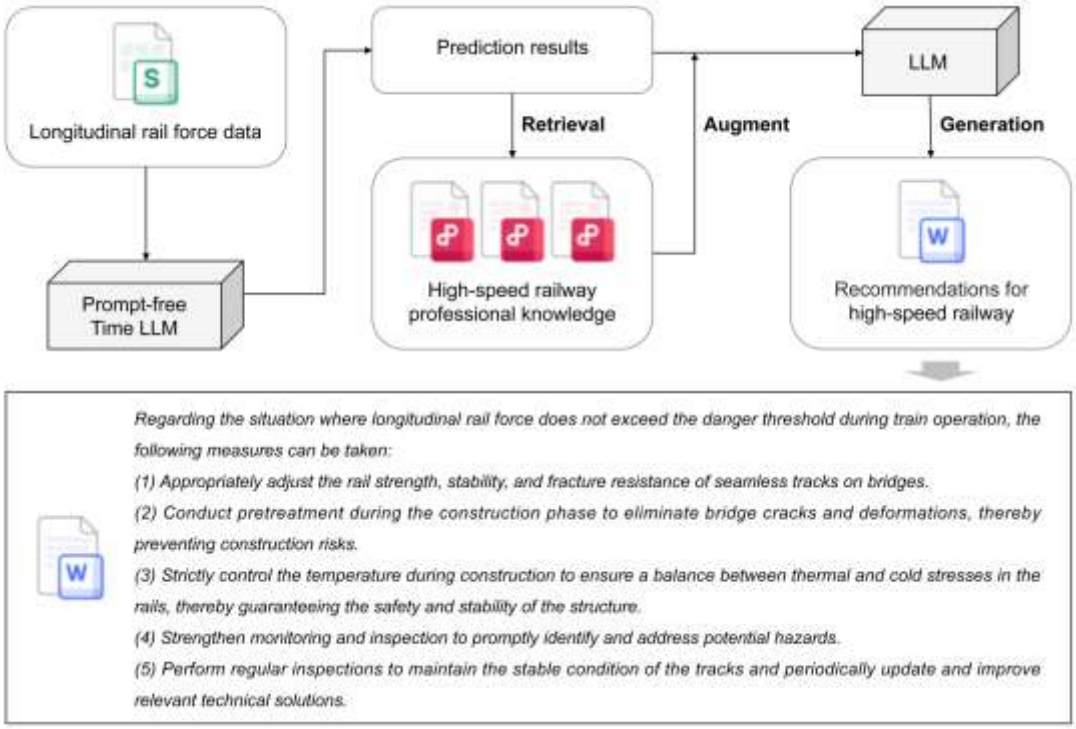


Fig. 7. Local-LLM based on retrieval-augmented generation.

Furthermore, Langchain-Chatchat enables professional question-answering capabilities grounded in localized domain knowledge of high-speed railways. Railway engineers can directly input their intended queries, which are then processed through retrieval-augmented generation to yield specialized responses. In addition, engineers can further expand the knowledge base to integrate interdisciplinary insights, thereby generating comprehensive answers that incorporate cross-domain expertise.

In summary, the Local-LLM serves dual functions: (1) processing uploaded high-speed railway monitoring data to perform longitudinal rail force prediction and output risk assessment, and (2) functioning as an expert system to address professional inquiries from engineers.

The predicted high-speed rail longitudinal rail force results in the dataset of this study were all within the safe range based on logical judgment. Based on this, the Local-LLM provided the following recommendations which can assist railway engineers:

Regarding the situation where longitudinal rail force does not exceed the danger threshold during train operation, the following measures can be taken:

(1)Appropriately adjust the rail strength, stability, and fracture resistance of seamless tracks on bridges.

(2)Conduct pretreatment during the construction phase to eliminate bridge cracks and deformations, thereby preventing construction risks.

(3)Strictly control the temperature during construction to ensure a balance between thermal and cold stresses in the rails, thereby guaranteeing the safety and stability of the structure.

(4)Strengthen monitoring and inspection to promptly identify and address potential hazards.

Perform regular inspections to maintain the stable condition of the tracks and periodically update and improve relevant technical solutions.

4. Conclusion

This study focuses on the prediction of high-speed railway longitudinal rail force, proposing a LLM framework that accomplishes two primary functions: (1) uploading high-speed railway monitoring data to achieve prediction and output risk assessment, and (2) serving as an expert system to address professional inquiries from engineers. For high-speed railway longitudinal rail force prediction, this research introduces the application of prompt-free Time-LLM [39], utilizing a 10-day data set as the training set to predict longitudinal rail force for the subsequent day. Experimental results demonstrate that:

(1)The prompt-free Time-LLM [39] achieves superior performance metrics, with an average R^2 of 0.932 and RMSE of 3.537 kN, outperforming both LSTM and RNN models, as well as the Time-LLM. The longitudinal rail force at different locations varies significantly. Different random seeds can also lead to differences in prediction effects, among which the standard deviation of prompt-free Time-LLM is the smallest.

(2)Under conditions of data deficiency, prompt-free Time-LLM [39] maintains the best predictive performance even when half of the data is intermittent missing. However, both prompt-free Time-LLM and Time-LLM fail to achieve effective stress prediction when half of the daily data is continuous missing. These findings highlight the significant impact of data quality on prediction accuracy and demonstrate that data deficiencies leading to reduced prediction precision cannot be compensated for through prompt engineering.

(3)The primary factors influencing future high-speed railway longitudinal rail force are predominantly recent stress conditions rather than long-term historical stress patterns. Furthermore, high-speed railway longitudinal rail force data exhibits strong trend characteristics. Consequently, in the context of longitudinal rail force monitoring, greater emphasis should be placed on the observation and analysis of recent stress conditions to ensure accurate prediction and effective maintenance strategies.

This research holds significant implications for high-speed railway engineering. By introducing a Local-LLM framework and the innovative prompt-free Time-LLM [39] approach, it provides a robust and efficient solution for predicting longitudinal rail force, a critical factor in ensuring railway safety and operational reliability. The model's superior performance, particularly its ability to maintain accuracy under partial data deficiency, offers a practical tool for real-time monitoring and risk assessment. This capability enables timely maintenance interventions, potentially preventing catastrophic failures and reducing operational downtime. Furthermore, the integration of expert system functionality allows for immediate professional guidance, enhancing decision-making processes for engineers. The study's findings underscore the importance of data quality in predictive accuracy, providing valuable insights for future data collection and management strategies in railway infrastructure monitoring. Overall, this research contributes to the advancement of intelligent monitoring systems in high-

speed railways, promoting safer and more efficient railway operations.

Although this study validates the effectiveness of the prompt-free Time-LLM [39] in longitudinal rail force forecasting, several limitations exist. The model is only verified on the turnout section of the Beijing-Shanghai High-Speed Railway, lacking validation on more lines and complex scenarios. The input only relies on historical rail force data without fusing multi-source information such as temperature and train parameters. Besides, the model performance degrades significantly under severe continuous data loss, and the knowledge base of Local-LLM cannot be updated dynamically.

Future work will build a multi-scene dataset to enhance generalization ability. A multi-modal Time-LLM will be designed by integrating temperature, train operation, and vibration data. Meanwhile, the dynamic knowledge base and self-learning mechanism will be developed. Finally, the lightweight model will be deployed on edge devices to realize real-time intelligent early warning and refined maintenance for high-speed railways.

Declaration of competing interest

The authors declare that they have no conflicts of interest in this work.

Acknowledgments

The authors would like to thank Dr. Xueyang Tang from Beijing Jiaotong University for his support of this study. DeepSeek was used for language polishing. This work was supported by the Natural Science Foundation of Beijing, China (L251029), the Fundamental Research Funds for the Central Universities (2025QYBS007), the National Natural Science Foundation of China (U24A20198).

References

[1] Gong, W., Akbar, M. F., Jawad, G. N., Mohamed, M. F. P., & Wahab, M. N. A. (2022). Nondestructive testing technologies for rail inspection: A review. Coatings, 12(11), 1790.

[2] Wang, P., Gao, Y., Yang, Y., Tian, G., Yao, E., & Wang, H. (2013). Experimental studies and new feature extractions of MBN for stress measurement on rail tracks. IEEE Transactions on Magnetics, 49(8), 4858-4864.

[3] Kanematsu, Y., & Matsui, M. (2023). Several tests for verification of material deterioration evaluation method of rail steel using X-ray diffraction analysis. Wear, 530, 205002.

[4] Bombarda, D., Vitetta, G. M., & Ferrante, G. (2021). Rail diagnostics based on ultrasonic guided waves: An overview. Applied Sciences, 11(3), 1071.

[5] Chang, Y., Li, N., Zhao, J., Wang, Y., & Yang, Z. (2022). A novel ultrasonic guided wave-based method for railway contact wire defect detection. IEEE Transactions on Instrumentation and Measurement, 71, 1-9.

[6] Milne, D., Masoudi, A., Ferro, E., Watson, G., & Le Pen, L. (2020). An analysis of railway track behaviour based on distributed optical fibre acoustic sensing. Mechanical

Systems and Signal Processing, 142, 106769.
[7] Jayawickrema, U. M. N., Herath, H. M. C. M., Hettiarachchi, N. K., Sooriyaarachchi, H. P., & Epaarachchi, J. A. (2022). Fibre-optic sensor and deep learning-based structural health monitoring systems for civil structures: A review. Measurement, 199, 111543.
[8] Liu, H., Yao, L., & Wu, H. (2024). Application of mobile laser measurement system in railway inspection. In Journal of Physics: Conference Series (Vol. 2770, No. 1, p. 012001).
[9] Wang, Y., Cai, X., Liu, W., Tang, X., Yao, Y., & Cai, X. (2025). Seismic-induced damage characteristics of high-speed railway ballastless tracks: A comprehensive nonlinear numerical simulation method. Engineering Structures, 343, 121248.
[10] Zhang, Y., Zhou, L., Mahunon, A. D., Zhang, G., Peng, X., Zhao, L., & Yuan, Y. (2021). Mechanical performance of a ballastless track system for the railway bridges of high-speed lines: Experimental and numerical study under thermal loading. Materials, 14(11), 2876.
[11] Jing, G., Siahkouhi, M., Qian, K., & Wang, S. (2021). Development of a field condition monitoring system in high speed railway turnout. Measurement, 169, 108358.
[12] Liu, G., Li, P., Wang, P., Liu, J., Xiao, J., Chen, R., & Wei, X. (2021). Study on structural health monitoring of vertical vibration of ballasted track in high-speed railway. Journal of Civil Structural Health Monitoring, 11(2), 451-463.
[13] Sun, J., Meli, E., Song, X., Chi, M., Jiao, W., & Jiang, Y. (2023). A novel measuring system for high-speed railway vehicles hunting monitoring able to predict wheelset motion and wheel/rail contact characteristics. Vehicle System Dynamics, 61(6), 1621-1643.
[14] Wang, Q. A., Zhang, C., Ma, Z. G., Huang, J., Ni, Y. Q., & Zhang, C. (2021). SHM deformation monitoring for high-speed rail track slabs and Bayesian change point detection for the measurements. Construction and Building Materials, 300, 124337.
[15] Lee, J., Jeong, S., Lee, J., Sim, S. H., Lee, K. C., & Lee, Y. J. (2022). Sensor data-based probabilistic monitoring of time-history deflections of railway bridges induced by high-speed trains. Structural Health Monitoring, 21(6), 2518-2530.
[16] Xiao, X., Shen, W., & He, X. (2021). Track irregularity monitoring on high-speed railway viaducts: a novel algorithm with unknown input condensation. Journal of Engineering Mechanics, 147(6), 04021029.
[17] Gonzalo, A. P., Horridge, R., Steele, H., Stewart, E., & Entezami, M. (2022). Review of data analytics for condition monitoring of railway track geometry. IEEE Transactions on Intelligent Transportation Systems, 23(12), 22737-22754.
[18] Li, Q., Gao, J., Beck, J. L., Lin, C., Huang, Y., & Li, H. (2024). Probabilistic outlier detection for robust regression modeling of structural response for high-speed railway track monitoring. Structural Health Monitoring, 23(2), 1280-1296.
[19] Luo, Q., Li, J., & Zhang, Y. (2022). Monitoring subsidence over the planned Jakarta–Bandung (Indonesia) high-speed railway using Sentinel-1 multi-temporal InSAR data. Remote Sensing, 14(17), 4138.
[20] Zhu, J., Hu, H., He, Z., Guo, X., & Pan, W. (2021). A power-quality monitoring and assessment system for high-speed railways based on train-network-data center

integration. Railway Engineering Science, 29(1), 30-41.
[21] Xu, W., Deng, Y., Zhang, B., Zhang, J., Peng, Z., Hou, B., & Duan, J. (2022). Crevice corrosion of U75V high-speed rail steel with varying crevice gap size by in-situ monitoring. Journal of Materials Research and Technology, 16, 1856-1874.
[22] Luo, Q., Fu, H., Liu, K., Wang, T., & Feng, G. (2021). Monitoring of train-induced responses at asphalt support layer of a high-speed ballasted track. Construction and Building Materials, 298, 123909.
[23] Xin, L., Mingzhou, B., Zijun, W., Pengxiang, L., Hai, S., & Ye, Z. (2021). Dynamic response and stability analysis of high-speed railway subgrade in karst areas. IEEE Access, 9, 129188-129206.
[24] Hoelzl, C., Dertimanis, V., Landgraf, M., Ancu, L., Zurkirchen, M., & Chatzi, E. (2022). On-board monitoring for smart assessment of railway infrastructure: A systematic review. The Rise of Smart Cities, 223-259.
[25] Jing, G., Qin, X., Wang, H., & Deng, C. (2022). Developments, challenges, and perspectives of railway inspection robots. Automation in Construction, 138, 104242.
[26] Sol-Sánchez, M., Castillo-Mingorance, J. M., Moreno-Navarro, F., & Rubio-Gámez, M. C. (2021). Smart rail pads for the continuous monitoring of sensored railway tracks: Sensors analysis. Automation in Construction, 132, 103950.
[27] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
[28] Ma, Z., & Gao, L. (2021). Predicting mechanical state of high-speed railway elevated station track system using a hybrid prediction model. KSCE Journal of Civil Engineering, 25(7), 2474-2486.
[29] Cai, X., Chang, W., Gao, L., & Zhou, C. (2021). Design and application of real-time monitoring system for service status of continuously welded turnout on the high-speed railway bridge. Journal of Transportation Safety & Security, 13(7), 735-758.
[30] Ling, Z., Hu, F., Liu, T., Jia, Z., & Han, Z. (2023). Hierarchical deep reinforcement learning for self-powered monitoring and communication integrated system in high-speed railway networks. IEEE Transactions on Intelligent Transportation Systems, 24(6), 6336-6349.
[31] Chen, S. X., Zhou, L., & Ni, Y. Q. (2022). Wheel condition assessment of high-speed trains under various operational conditions using semi-supervised adversarial domain adaptation. Mechanical Systems and Signal Processing, 170, 108853.
[32] Cai, X., Tang, X., Pan, S., Wang, Y., Yan, H., Ren, Y., Chen, N., & Hou, Y. (2024). Intelligent recognition of defects in high-speed railway slab track with limited dataset. Computer-Aided Civil and Infrastructure Engineering, 39(6), 911-928.
[33] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.
[34] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.
[35] Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274.

[36] Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. (2023). Large language models in medicine. Nature Medicine, 29(8), 1930-1940.
[37] Chen, W., Yan-yi, L., Tie-zheng, G., Da-peng, L., Tao, H., Zhi, L., ... & Ying-you, W. (2024). Systems engineering issues for industry applications of large language model. Applied Soft Computing, 151, 111165.
[38] Wu, T., & Ling, Q. (2024). STELLM: Spatio-temporal enhanced pre-trained large language model for wind speed forecasting. Applied Energy, 375, 124034.
[39] Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J. Y., Shi, X., ... & Wen, Q. (2023). Time-LLM: Time series forecasting by reprogramming large language models. arXiv preprint arXiv:2310.01728.
[40] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533-536.
[41] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.
[42] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI Blog, 1(8).
[43] Liu, Q., Song, J., Huang, Z., Zhang, Y., glide-the, & liunux4odoo. (2024). Langchain-chatchat [Computer software]. GitHub. https://github.com/chatchat-space/Langchain-Chatchat.
[44] Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Bhargava, P., Bhosale, S., et al. (2023). Llama 2: Open foundation and fine-tuned chat models. arXiv Preprint arXiv:2307.09288.
[45] Wang, Y., Cai, X., Tang, X., Pan, S., Wang, Y., Yan, H., Ren, Y., & Hou, Y. (2024). HSRA-Net: Intelligent Detection Network of Anomaly Monitoring Data in High-Speed Railway. IEEE Transactions on Intelligent Transportation Systems.