This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TII.2024.3452207,

Separable Convolutional Network-based Fault Diagnosis for High-Speed Train: A Gossip Strategy-based Optimization Approach

Yihao Xue, *Student Member, IEEE*, Rui Yang, *Senior Member, IEEE*, Xiaohan Chen, *Student Member, IEEE*, Baoye Song, Zidong Wang, *Fellow, IEEE*

***Abstract*— With the rapid development of high-speed train, health monitoring of high-speed train traction power system has gradually become a popular research topic. The traction asynchronous motor, as a key component in the traction power systems, greatly affects the reliability, stability, and safety of high-speed train operation. Normally, when faults occur, the train needs to immediately slow down or even stop to avoid unimaginable losses, resulting in limited fault data. Traditional data-driven fault diagnosis methods may face the local optimum problem during the optimization process when training samples are insufficient. In this study, a novel gossip strategy-based fault diagnosis method is proposed to prevent the local optimum problem, thus improving fault diagnosis performance. The proposed gossip strategy-based fault diagnosis method is validated on the hardware-in-the-loop (HIL) high-speed train traction control system simulation platform, and the experimental results unequivocally show that the proposed method outperforms other well-known methods.**

***Index Terms*— high-speed train, fault diagnosis, local optimum, gossip strategy, neural network.**

I. Introduction

HIGH-SPEED train has the advantages of punctuality, comfort, and convenience, and has become a mainstream transportation choice in recent years. Due to its widespread adoption, the health monitoring of high-speed trains is paramount to ensure reliability and stability during operation [26]. The traction system, a crucial component of high-speed trains, can provide power during operation. Among the components of the traction system, the asynchronous motor plays a vital role in determining the power, energy consumption, and control characteristics of high-speed train [39]. Once the traction asynchronous motor fails, the power supply to the system is immediately interrupted, leading to disruptions in normal railway operations. Therefore, the health monitoring and fault diagnosis of traction asynchronous motors are of utmost importance for ensuring the reliable operation of high-speed train.

In recent years, researchers have shown widespread interest in data-driven fault diagnosis methods [8], [38]. These approaches can involve establishing diagnostic models based on historical data without the need to construct complex physical models [21], [48]. Therefore, data-driven methods are widely applied in various fields such as state estimation [7], [30], [43], consensus control [4], [11], [13], object detection [10], [36], [46], optimal control [20], [22], [37], and fault diagnosis [2], [15], [34]. As one of the main research directions of data-driven methods, deep learning technology incorporates feature learning into the model-building process, enabling automatic learning of pattern features from samples [29], [35]. When sufficient fault data is available, these methods can effectively extract deep fault features and perform accurate fault diagnosis [26], [27]. However, for high-speed train fault diagnosis, obtaining an ample amount of fault data is challenging. In the event of a fault in the traction system of a high-speed train, immediate deceleration or even a complete stop is necessary to prevent substantial losses. Consequently, only a restricted amount of fault data can be recorded and utilized for fault diagnosis, posing challenges for existing deep learning-based methods in high-speed train fault diagnosis [5], [33].

Considering the scarcity of data in high-speed train fault diagnosis, accurately estimating complex model parameters poses a challenge, restricting the overall model performance and impeding the comprehensive capture of intricate structures and representations within the fault data. Such a challenge often manifests as the model getting trapped in local optimum within the parameter space [45]. As a result, facing the frequent challenge of local optimum in existing fault diagnosis methods, especially in scenarios with limited fault data, current research has proposed various solutions. These efforts

This work is partially supported by the National Natural Science Foundation of China (62233012), the Jiangsu Provincial Qinglan Project (2021), the Research Development Fund of XJTLU (RDF-20-01-18), XJTLU Research Enhancement Fund (REF-23-01-008) and the Suzhou Science and Technology Programme (SYG202106).
Yihao Xue and Xiaohan Chen are with the School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China, 215123, and also with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, L69 3GJ, United Kingdom (email: Yihao.Xue21@student.xjtlu.edu.cn; Xiaohan.Chen20@student.xjtlu.edu.cn).
Rui Yang is with the School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, 215123, China (email: R.Yang@xjtlu.edu.cn).
Baoye Song is with the College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, China (email: Songbaoye@sdust.edu.cn).
Zidong Wang is with the Department of Computer Science, Brunel University London, Uxbridge, Middlesex UB8 3PH, United Kingdom (email: Zidong.Wang@brunel.ac.uk).
Corresponding author: Rui Yang

primarily focus on two key aspects: optimization algorithm improvement [9] and lightweight model design [40].

To address the potential local optimum problem in high-speed train fault diagnosis, researchers have made improvements to the traditional deep learning methods by focusing on enhancing optimization algorithms and designing lightweight models [9], [40]. Regarding the enhancement of optimization algorithms, the emphasis is typically on two aspects: (1) Incorporating the momentum term into the gradient descent algorithms to effectively prevent the drawback of easily getting trapped in local optimum, thereby improving the model stability during convergence process [23]; (2) Implementing adaptive learning rate methods, where small (large) learning rate is applied to update low-frequency (high-frequency) parameters [45]. However, relying solely on momentum terms and adaptive learning rates may not always be sufficient to ensure that the models will not trap in local optimum during convergence process, leaving the possibility for further improvement in existing methods.

Dedicated to alleviating the local optimum issue, researchers have proposed numerous lightweight models for high-speed train fault diagnosis with limited samples [6], [38]. Lightweight models aim to effectively reduce model size and computational parameters, extracting common fault features rather than specific ones, thus avoiding local optimum caused by limited data [14]. However, existing studies often utilize pruning algorithms or employ concise model structures to reduce the computational parameters, which may result in incomplete or insufficient in-depth information due to structural constraints, subsequently affecting the diagnostic accuracy. Noted that the limitations of existing methods for avoiding local optimum problem are listed in Table I.

TABLE I
LIMITATIONS OF EXISTING METHODS IN AVOIDING LOCAL OPTIMUM.

Existing methods	Literature	Limitations
Momentum term	[9], [18], [23]	Risk of oscillation or divergence.
Adaptive learning rate	[9], [16], [45]	Persistence of local optimum problem.
Pruning algorithm	[6], [12], [42]	Potential loss of beneficial neuron.
Concise model structure	[14], [38], [40]	Incomplete in-depth feature.

Based on the literature analysis above, although existing research on optimization algorithm enhancement and lightweight model design can address the potential local optimum issue in high-speed train fault diagnosis, there is possibility for further improvement. An intuitive idea is to propose an efficient optimization algorithm to enhance the model convergence performance and introduce a parameter-limited deep neural network model to ensure the model's proficiency in in-depth feature extraction. In this way, the improved fault diagnosis method can offer the following advantages: (1) Significantly avoiding local optimum during the optimization process, thereby promoting model convergence; (2) Effectively extracting in-depth fault features and circumventing local optimum due to the lightweight network structure, thereby enhancing diagnostic performance.

Drawing on the preceding discussion, the objective of this study is to enhance the efficacy of high-speed train fault diagnosis by proposing a novel neural network optimization algorithm and introducing a lightweight fault diagnosis method. Specifically, leveraging the benefits of the gossip strategy in knowledge sharing and consensus building, which can assist the model in exploring various regions of high-dimensional space during the convergence process to circumvent local optimum and converge towards the optimal global solution. Therefore, a novel gossip strategy-based optimization method is proposed in this study to steer clear of suboptimal solutions that may emerge during the convergence process. Moreover, capitalizing on the advantages of separable convolution and long short-term memory (LSTM) in reducing the model parameters while maintaining excellent feature extraction capabilities [24], [38], this study devises an efficient separable convolution LSTM (SC-LSTM) model. In comparison to other sophisticated neural network models, the constructed SC-LSTM can exhibit exceptional proficiency in capturing temporal dependencies in lengthy sequences and has remarkable generalization and expression abilities in environments with limited fault data.

In summary, the proposed intelligent gossip strategy-based SC-LSTM method can adeptly extract fault features with limited model parameters while averting the potential local optimum problem in high-speed train fault diagnosis. The key contributions of this study include:

1) A novel gossip strategy-based optimization algorithm is proposed to address the local optimum challenge during the model convergence.
2) An efficient gossip strategy-based SC-LSTM method is proposed for intelligent high-speed train fault diagnosis, aiming to capture temporal dependencies of extensive sequences and extract in-depth fault features with limited model parameters, enabling the mitigation of local optimum issue.
3) The proposed gossip strategy-based SC-LSTM method is successfully applied to fault diagnosis in high-speed train traction motors, and experimental results demonstrate its superiority over six existing fault diagnosis methods. Additionally, the proposed gossip strategy-based optimization algorithm demonstrates satisfactory results under different initialization conditions compared to several well-known optimization algorithms.

The remaining parts of this paper can be organized in three sections. Section II provides a technical description and explanation of the proposed gossip strategy-based fault diagnosis method. In Section III, the proposed method is compared with six existing fault diagnosis methods on the hardware-in-the-loop (HIL) high-speed train traction control system simulation platform. Finally, Section IV summarizes the conducted research and outlines future works.

II. METHODOLOGY

A. Gossip Strategy-based Optimization Algorithm

The origin of the gossip strategy can be traced back to the research on distributed systems [1]. One significant challenge in distributed systems is to ensure that all nodes in the network have consistent information, especially when the network is dynamic or subject to failures. The gossip strategy can facilitate direct information sharing among nodes without relying

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/TII.2024.3452207, IEEE Transactions on Industrial Informatics

on a centralized coordinator or broadcasting information to all nodes. This decentralized information propagation enhances system robustness and fault-tolerance. In the event of a node failure, information can still propagate through the remaining nodes, swiftly updating information in a dynamic system.

Building on the aforementioned principles, this study integrates the gossip strategy into the convergence process of neural networks and proposes a novel gossip strategy-based optimization algorithm tailored for high-speed train fault diagnosis. This integration allows the model to circumvent local optimum through information exchange, thereby enhancing the convergence performance of traditional neural network optimization algorithms. Concretely, a designated number of recorders are incorporated during the convergence process, and in each iteration, these recorders probabilistically save the current model parameters. During the model convergence process, if the loss value of the model ceases to decrease after a predetermined number of iterations, all recorders engage in a weighted average computation between the recorded model parameters and the current model parameters. In instances where the model converges to a local optimum, the information exchange and parameter combination mechanisms among recorders contribute to bolstering the global exploration capability of the optimization process, facilitating escape from the local optimum. The proposed gossip strategy-based optimization algorithm follows a structured four-step process, as illustrated in Fig. 1, including local update, information exchange, parameter combination, and collective update.

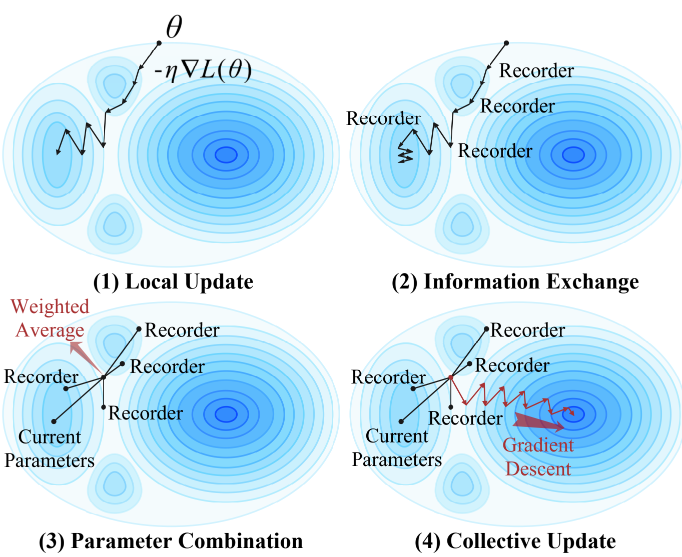

Fig. 1. Flowchart of the proposed gossip strategy-based neural network optimization algorithm.

1) Local Update: In the local update phase, the SoftMax cross-entropy function is adopted to classify multi-class faults. For the training set $\left\{\left(x^{(1)}, y^{(1)}\right), \ldots, \left(x^{(\Im)}, y^{(\Im)}\right)\right\}$, we have $y^{(i)} \in \{1, 2, \ldots, \varkappa\}$ with $\varkappa$ classes. For a given test sample $x^{(i)}$, a hypothesis function needs to be adopted to estimate the probability value $p(y = j|x^{(i)})$ for each category j, and the hypothesis function $h_\theta(x)$ can be represented as follows:

$$h_\theta\left(x^{(i)}\right) = \begin{bmatrix} p\left(y^{(i)} = 1|x^{(i)}; \theta\right) \\ p\left(y^{(i)} = 2|x^{(i)}; \theta\right) \\ \vdots \\ p\left(y^{(i)} = \varkappa|x^{(i)}; \theta\right) \end{bmatrix} = \frac{1}{\sum_{l=1}^{\varkappa} e^{\theta_l^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_\varkappa^T x^{(i)}} \end{bmatrix} \quad (1)$$

where $\theta = [\theta_1, \theta_2, \ldots, \theta_\varkappa]$ represents the set of parameters associated with each class in the multi-class classification model. The cost function can be represented as follows:

$$L(\theta) = -\frac{1}{\Im}\left[\sum_{i=1}^{\Im}\sum_{j=1}^{\varkappa} \mathbf{1}\{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{\varkappa} e^{\theta_l^T x^{(i)}}}\right] \quad (2)$$

where $j = 1, 2, \ldots, \varkappa$, and $\mathbf{1}$ represents a logic function outputting 1 for a true condition and 0 otherwise.

To address the minimization problem of the cost function $L(\theta)$, firstly, the cost function $L(\theta)$ needs to be simplified into the following form:

$$\begin{aligned} L(\theta) = &-\frac{1}{\Im}\left[\sum_{i=1}^{\Im}\left(\sum_{j=1}^{\varkappa} \mathbf{1}\{y^{(i)} = j\} \cdot \theta_j^T x^{(i)}\right.\right. \\ &\left.\left.-\sum_{j=1}^{\varkappa} \mathbf{1}\{y^{(i)} = j\} \log \sum_{l=1}^{\varkappa} e^{\theta_l^T x^{(i)}}\right)\right] \\ = &-\frac{1}{\Im}\left[\sum_{i=1}^{\Im}\left(\sum_{j=1}^{\varkappa} \mathbf{1}\{y^{(i)} = j\} \cdot \theta_j^T x^{(i)} - \log \sum_{l=1}^{\varkappa} e^{\theta_l^T x^{(i)}}\right)\right] \end{aligned} \quad (3)$$

Then the partial derivative of $L(\theta)$ is calculated as below:

$$\begin{aligned} &\frac{\partial L(\theta)}{\partial \theta_j} \\ &= -\frac{1}{\Im}\left[\sum_{i=1}^{\Im}\left(\frac{\partial \sum_{j=1}^{\varkappa} \mathbf{1}\left\{y^{(i)} = j\right\} \cdot \theta_j^T x^{(i)}}{\partial \theta_j} - \frac{\partial \log \sum_{l=1}^{\varkappa} e^{\theta_l^T x^{(i)}}}{\partial \theta_j}\right)\right] \\ &= -\frac{1}{\Im}\left[\sum_{i=1}^{\Im}\left(\sum_{j=1}^{\varkappa} \mathbf{1}\{y^{(i)} = j\} x^{(i)} - \frac{\partial \sum_{l=1}^{\varkappa} e^{\theta_l^T x^{(i)}}}{\partial \theta_j} \cdot \frac{1}{\sum_{l=1}^{\varkappa} e^{\theta_l^T x^{(i)}}}\right)\right] \\ &= -\frac{1}{\Im}\left[\sum_{i=1}^{\Im}\left(\sum_{j=1}^{\varkappa} \mathbf{1}\{y^{(i)} = j\} x^{(i)} - \frac{e^{\theta_j^T x^{(i)}} \partial\left(\theta_j^T x^{(i)}\right)/\partial \theta_j}{\sum_{l=1}^{\varkappa} e^{\theta_l^T x^{(i)}}}\right)\right] \\ &= -\frac{1}{\Im}\left[\sum_{i=1}^{\Im}\left(\sum_{j=1}^{\varkappa} \mathbf{1}\{y^{(i)} = j\} x^{(i)} - \frac{e^{\theta_j^T x^{(i)}} x^{(i)}}{\sum_{l=1}^{\varkappa} e^{\theta_l^T x^{(i)}}}\right)\right] \\ &= -\frac{1}{\Im}\sum_{i=1}^{\Im}\left[x^{(i)}\left(\sum_{j=1}^{\varkappa} \mathbf{1}\{y^{(i)} = j\} - p\left(y^{(i)} = j|x^{(i)}; \theta\right)\right)\right] \end{aligned} \quad (4)$$

Finally, the gradient function is obtained as below:

$$\nabla L(\theta) = \left[\frac{\partial L(\theta)}{\partial \theta_1}, \frac{\partial L(\theta)}{\partial \theta_2}, \ldots, \frac{\partial L(\theta)}{\partial \theta_\varkappa}\right] \quad (5)$$

The obtained gradient function $\nabla L(\theta)$ is plugged into the stochastic gradient descent algorithm [31] to minimize $L(\theta)$. Specifically, assuming that the model parameters recorded by the recorders at the t-th iteration during model training process are represented as $\theta_{(t)}$, in each iteration of the neural network over the specified subset of training data $(X_i; y_i)$, the following update needs to be performed:

$$\theta(t) = \theta(t-1) - \eta \nabla L(\theta(t-1)) \quad (6)$$

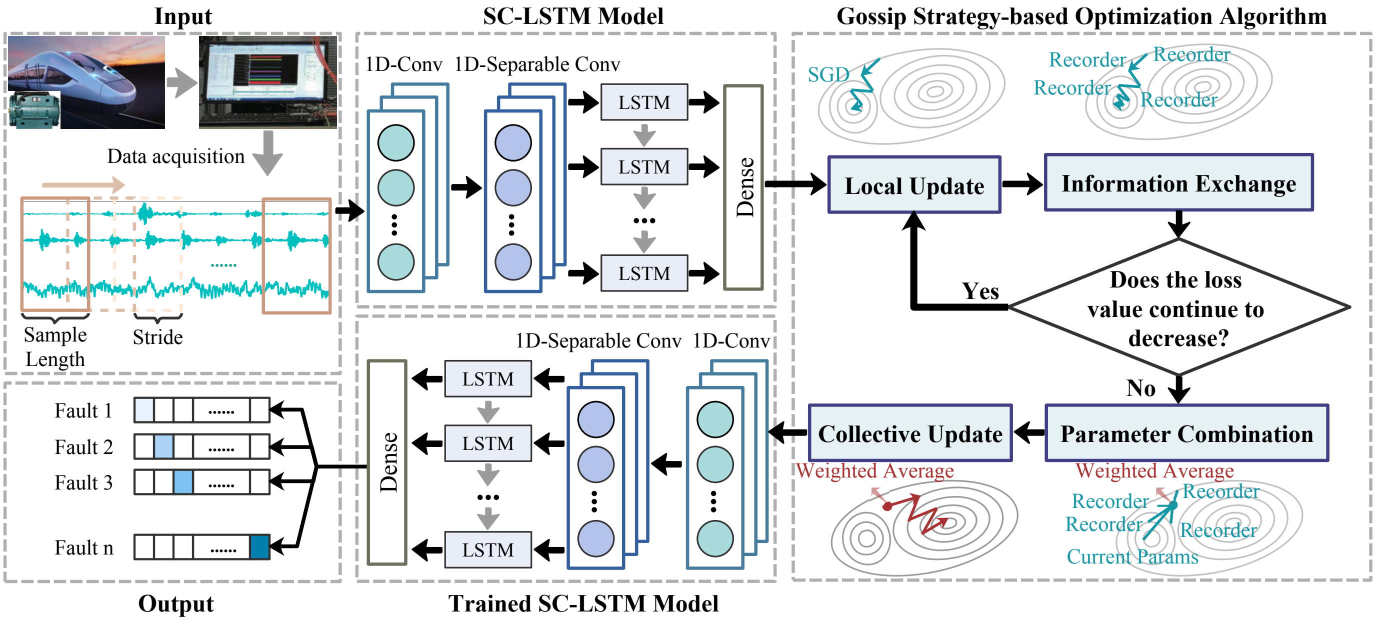

Fig. 2. Structure diagram of the proposed gossip strategy-based SC-LSTM fault diagnosis model.

where η represents the learning rate, and $\nabla L(\theta(t-1))$ signifies the gradient of the loss function L concerning the parameters $\theta(t-1)$ computed on a mini-batch training data $(X_i; y_i)$.

2) Information Exchange: In this step, a series of recorders are introduced to store the model parameters at different iteration periods. During the model convergence period, there exists a slight probability that the model parameters of the current iteration will be saved to a random recorder. This implies that there is a certain probability of sharing between the model parameters of different iteration periods and the recorders. Assuming there are k recorders participating in the information exchange step, the parameters received by the i-th recorder can be signified as $\theta^i(t)$ $(i \in [1, 2, \dots, k])$, i.e. the model parameters saved by the recorder at the t-th iteration.

3) Parameter Combination: If the loss value cannot be reduced after a certain number of iterations during the model convergence, it is possible that the model has been trapped in local optimum. In this way, the recorders that stored previous model parameters need to engage in combination operation with the current model parameters, thereby assisting the model escape from potential local optimum. In other words, the model needs to combine the parameters of current iteration with the parameters received by the recorders at different iterations. In this study, a weighted average of the current model parameters and the model parameters saved by the recorders is performed. Let β_i denotes the weight associated with the parameters stored by the i-th recorder, the combined parameters $\theta^c(t)$ for recorder i can be computed as below:

$$\theta^c(t) = \alpha \cdot \theta(t) + \sum_{i=1}^{k} \beta_i \cdot \theta^i(t) \quad (7)$$

where α is a weighting factor that determines the proportion of the current model parameters. The weight β_i can represent the proportion of model parameters saved by the recorders, which is predefined or dynamically determined based on factors such as information exchange frequency, recorder performance, or other heuristics. Through parameter combination, it is able to integrate the parameter information saved by the recorders with the parameter information of the current iteration. If the model of current state is in a local optimum, the parameter combination operation can assist the model in escaping from the local optimum state.

4) Collective Update: After the parameter combination step, a set of updated model parameters can be obtained via the weighted average of the parameters from the current iteration and the parameters saved by the recorders. In the collective update step, the model is updated via gradient descent based on the updated model parameters, with the collective update process expressed as follows:

$$\theta(t+1) = \theta^c(t) - \eta \nabla L(\theta^c(t)) \quad (8)$$

B. Fault Diagnosis Procedure based on Gossip Strategy-based SC-LSTM Method

This paper proposes a novel lightweight gossip strategy-based SC-LSTM method for high-speed train fault diagnosis. By incorporating a lightweight model structure and an innovative optimization algorithm, the proposed method can mitigate the local optimum challenge and improve diagnostic performance. The process of the proposed method is depicted in Fig. 2, with detailed model information exhibited in Table II. The model takes the high-speed train fault signals as the input data, dividing the collected signals into multiple sample segments. The lightweight SC-LSTM model comprises convolutional layers, separable convolutional layers, and LSTM layers. Following the model feature extraction, the proposed gossip strategy-based optimization algorithm is employed for iterative model convergence. Ultimately, the trained model is utilized for accurate and efficient fault diagnosis.

In the proposed fault diagnosis method exhibited in Fig. 2 and Table II, an efficient lightweight SC-LSTM model is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/TII.2024.3452207, IEEE Transactions on Industrial Informatics

TABLE II
DETAILED INFORMATION OF THE SC-LSTM.

Layer	Filters (Units)	Kernel/ Stride	Activation Function	Input	Output
Input	/	/	/	$(N,1)$	$(N,1)$
Conv1D	32	$64/2$	ReLU	$(N,1)$	$(N/2,32)$
Batch Norm	/	/	/	$(N/2,32)$	$(N/2,32)$
Conv1D	32	$64/2$	ReLU	$(N/2,32)$	$(N/4,32)$
Batch Norm	/	/	/	$(N/4,32)$	$(N/4,32)$
Depthwise Conv1D	32	$8/1$	ReLU	$(N/4,32)$	$(N/4,32)$
Pointwise Conv1D	32	$1/1$	ReLU	$(N/4,32)$	$(N/4,32)$
Batch Norm	/	/	/	$(N/4,32)$	$(N/4,32)$
Depthwise Conv1D	32	$8/1$	ReLU	$(N/4,32)$	$(N/4,32)$
Pointwise Conv1D	32	$1/1$	ReLU	$(N/4,32)$	$(N/4,32)$
Batch Norm	/	/	/	$(N/4,32)$	$(N/4,32)$
Flatten	/	/	/	$(N/4,32)$	$(8N)$
Reshape	/	/	/	$(8N)$	$(N/16,128)$
LSTM	128	/	/	$(N/16,128)$	$(N/16,128)$
LSTM	64	/	/	$(N/16,128)$	$(N/16)$
Dense	/	/	SoftMax	$(N/16)$	n

introduced to achieve accurate fault diagnosis for high-speed trains and avoid the potential local optimum issue. Specifically, considering the effectiveness of conventional convolutional layers in fault feature extraction, they are utilized to capture the fault patterns and general characteristics of the input samples. Subsequently, the separable convolutional layers are set after the convolutional layers to delve into profound fault features while significantly reducing the parameters involved in convolutional operations. Table II delineates that the separable convolution comprises two parts: depthwise convolution and pointwise convolution. The depthwise convolution operates independently on feature vectors within each channel, capturing in-depth features of the fault signals; simultaneously, the pointwise convolution performs weighted operations on each channel's feature vectors to establish relationships between channels and obtain the final feature representation [38]. Additionally, the LSTM layers are employed to leverage long-term memory and low computational parameters, capturing temporal fault characteristics, and effectively addressing issues like gradient vanishing or exploding [24].

To enhance clarity in illustrating the fault diagnosis process, Algorithm 1 presents the pseudocode of the proposed fault diagnosis method. The proposed gossip strategy-based fault diagnosis method mainly consists of five steps:

1) *Data Acquisition and Preprocessing*: collect the relevant operational signal data (such as current, voltage, speed, etc.) from the high-speed train traction motor, divide the collected long sequential signal data into sample segments, and conduct normalization.
2) *Model Initialization*: initialize the weights and biases of the introduced SC-LSTM model.
3) *Model Training*: randomly shuffle the operational signal segments and partition them into training and test sets. In each iteration, select a specified subset from the training set for model training and update the network parameters utilizing the backpropagation algorithm.
4) *Iterative Optimization*: utilize the proposed gossip strategy-based optimization algorithm to enhance the effectiveness of gradient descent and mitigate the local optimum problem.
5) *Model Validation*: utilize the SoftMax function to compute the distribution probability of each test sample across different fault categories. The fault prediction result is determined by selecting the category with the highest probability.
6) *Fault Diagnosis*: upon receiving new operational data, the trained model is capable of conducting fault diagnosis. The model can provide output information regarding the state of the high-speed train traction system, encompassing the identification of faults and the categorization of specific fault types.

Algorithm 1 Gossip strategy-based fault diagnosis method

Input data: $D = \left\{x^{(i)}, y^{(i)}\right\}_{i=1}^{\Im}$
Output: The optimal fault prediction $\left\{\widehat{y}^{(j)}\right\}_{j=1}^{\Im}$

Initialize the model
for each *epoch* **do**
  for each *batch* **do**
    1. Extract the in-depth and time-series features of $x^{(i)}$ by separable convolution and LSTM;
    2. Compute cost function $L(\theta)$ by (1) and (2);
    3. Obtain the gradient $\nabla L(\theta)$ by (3) to (5), and conduct local update of parameters θ via (6);
    4. Save the current model parameters $\theta(t)$ with a tiny probability to a random recorder for information exchange;
    5. Perform parameter combination by (7) if the loss value no longer decreases;
    6. Continue to update the combined model parameters $\theta^c(t)$ via (8).
  end for
  Log the current classification result $\left\{\widehat{y}^{(j)}\right\}_{j=1}^{\Im}$
end for
return The optimal fault prediction $\left\{\widehat{y}^{(j)}\right\}_{j=1}^{\Im}$

III. EXPERIMENTAL RESULTS AND ANALYSIS

The proposed gossip strategy-based fault diagnosis method is validated on the HIL high-speed train traction control system simulation platform as shown in Fig. 3. In this paper, the proposed method is compared to other recently published methods on the same simulation platform. The models used in this study are implemented using TensorFlow and Keras, trained on a server featuring a Xeon(R) Intel(R) CPU E5-2678v3@2.50GHz, 64-GB main memory, and an NVIDIA Titan RTX GPU.

A. Description of the HIL High-Speed Train Traction Control System Simulation Platform

The current signal data are sourced from the HIL high-speed train traction control system simulation platform, jointly established by Central South University and Zhuzhou Electric Locomotive Research Institute, as shown in Fig. 3. This simulation platform comprises five main components: a signal conditioner, a personal computer (PC), a traction control unit (TCU), a power source and a dSPACE real time simulator

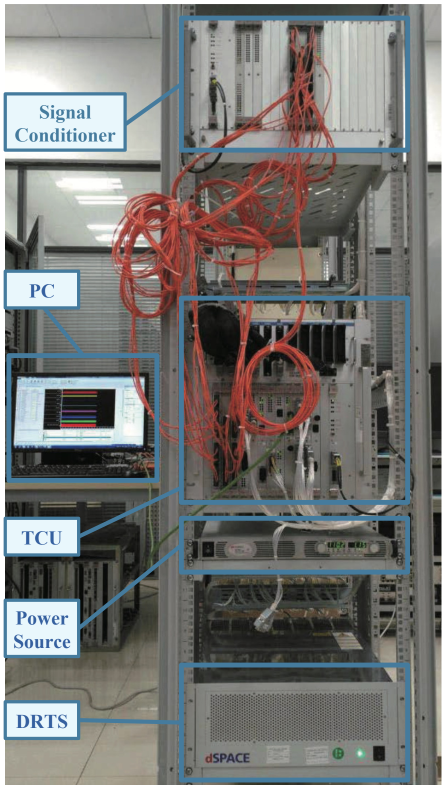

Fig. 3. HIL high-speed train traction control system simulation platform.

(DRTS). The dSPACE simulator is utilized to implement high-speed computation of the simulation platform under both normal and fault states, while the TCU is equipped with a traction asynchronous motor control program and hardware protection measures. To inject faults in the HIL setup, modifications to the relevant system settings can be made, following the traction and driving control system fault injection benchmark platform [41]. The collected dataset encompasses pre-set current signals for three fault types: *rotor broken bars (RBB)*, *inter-turn short circuit (ISC)*, and *air gap eccentricity (AGE)*. These current signals are acquired through sensors monitoring the rectifier AC-side current, under specific working condition (*280 km/h*) and severity levels (*minor*, *moderate*, and *serious*), with a sampling frequency of 2.5 kHz. Consequently, there are ten operation states, comprising one normal state and nine fault states, each with specific information outlined in Table III, with visualization exhibited in Fig. 4.

TABLE III
FAULT DETAILS OF HIL HIGH-SPEED TRAIN SIMULATION PLATFORM.

Label	Fault Type	Severity Level	Fault Location	Working Condition
(a)	Normal	None	Traction Asynchronous Motor	280 km/h
(b)	RBB	Minor	Traction Asynchronous Motor	280 km/h
(c)	RBB	Moderate	Traction Asynchronous Motor	280 km/h
(d)	RBB	Serious	Traction Asynchronous Motor	280 km/h
(e)	ISC	Minor	Traction Asynchronous Motor	280 km/h
(f)	ISC	Moderate	Traction Asynchronous Motor	280 km/h
(g)	ISC	Serious	Traction Asynchronous Motor	280 km/h
(h)	AGE	Minor	Traction Asynchronous Motor	280 km/h
(i)	AGE	Moderate	Traction Asynchronous Motor	280 km/h
(j)	AGE	Serious	Traction Asynchronous Motor	280 km/h

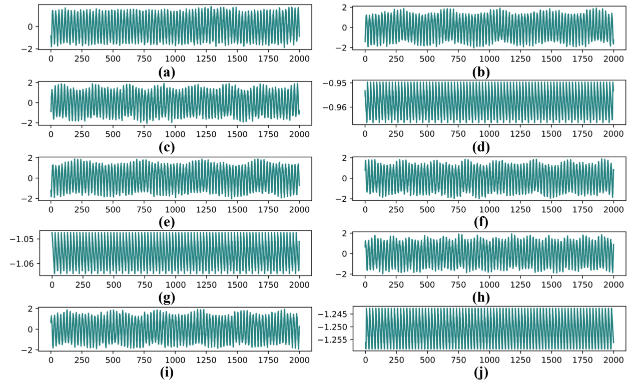

Fig. 4. Normalized rectifier AC-side current signals for ten operation states of the HIL high-speed train traction control system simulation platform.

B. Performance Comparison with Existing Methods

In this study, we validate the proposed high-speed train fault diagnosis approach by benchmarking it against six recently published deep learning-based methods: 1-dimensional CNN (1DCNN) [17], CNN-LSTM (CLSTM) [3], deep CNN (DCNN) [28], gated recurrent unit-multilayer perceptron (GRU-M) [47], 6-layer residual neural network (ResNet06) [44], and 1-dimensional separable convolutional neural network (Sep-CNN) [14]. To underscore the effectiveness of the proposed SC-LSTM method, we also explore lightweight alternatives, including CLSTM, ResNet06, and Sep-CNN. These lightweight models aim to maintain diagnostic accuracy while reducing model complexity and computational demands. Additionally, to comprehensively explicate the superiority of the proposed method over other methods, the ablation experiments are conducted across various fault diagnosis methods. Specifically, the gossip strategy-based optimization algorithm is selectively applied to enhance the optimization process of each model. Furthermore, ten repeated experiments are undertaken to assess the effectiveness of various methods, and the experimental results are exhibited in Fig. 5 and Table IV, comprising the best accuracy, worst accuracy, average accuracy and the model trainable parameters. It is noteworthy that the ten repeated experiments use the same number of training samples but different randomly selected subsets of

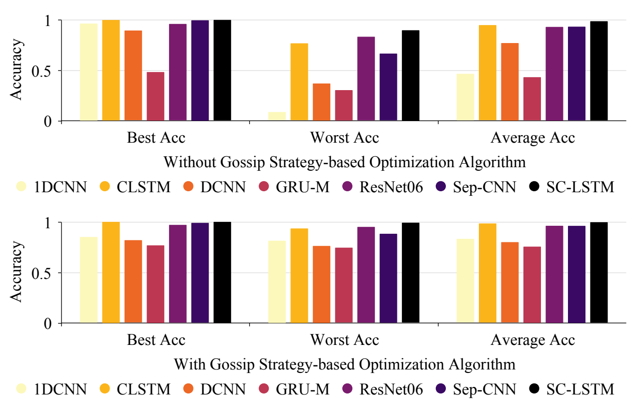

Fig. 5. Performance of different methods on HIL high-speed train traction control system simulation platform.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/TII.2024.3452207, IEEE Transactions on Industrial Informatics

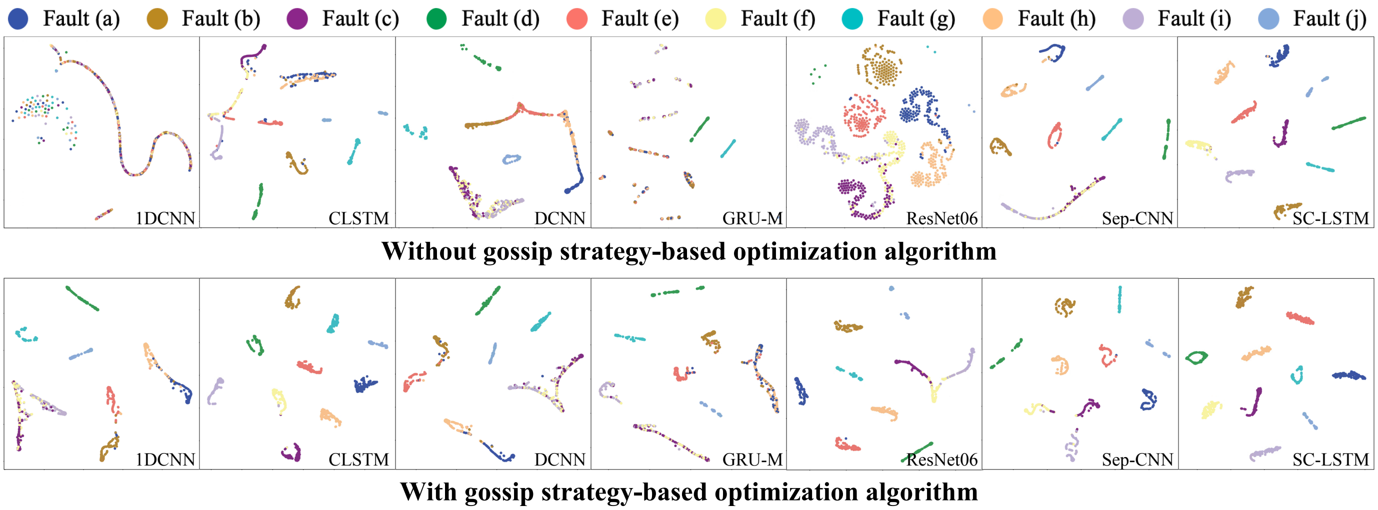

Fig. 6. Visualization of features extracted by different methods based on t-SNE.

TABLE IV
PERFORMANCE OF DIFFERENT METHODS ON HIL HIGH-SPEED TRAIN TRACTION CONTROL SYSTEM SIMULATION PLATFORM.

Methods	Best Acc	Worst Acc	Average Acc	Model Params
Without gossip strategy-based optimization algorithm				
1DCNN [17]	96.10%	8.35%	46.26%	9,004,042
CLSTM [3]	99.64%	76.50%	94.58%	861,922
DCNN [28]	89.20%	36.66%	76.76%	3,162,894
GRU-M [47]	48.00%	30.04%	42.91%	14,712,846
ResNet06 [44]	95.74%	83.03%	92.78%	1,074,986
Sep-CNN [14]	99.36%	66.33%	93.03%	**20,891**
SC-LSTM	**99.73%**	**89.47%**	**98.43%**	203,274
With gossip strategy-based optimization algorithm				
1DCNN [17]	85.03%	81.31%	83.18%	9,004,042
CLSTM [3]	**100.00%**	93.47%	98.45%	861,922
DCNN [28]	81.85%	76.13%	79.89%	3,162,894
GRU-M [47]	76.68%	74.41%	75.44%	14,712,846
ResNet06 [44]	96.64%	92.38%	95.41%	1,074,986
Sep-CNN [14]	99.00%	88.20%	96.06%	**20,891**
SC-LSTM	**100.00%**	**99.18%**	**99.66%**	203,274

TABLE V
PERFORMANCE ENHANCEMENT BY INCORPORATING GOSSIP STRATEGY-BASED OPTIMIZATION IN DIFFERENT METHODS.

Methods	Best Acc	Worst Acc	Average Acc
1DCNN [17]	−11.07%	72.96%	36.92%
CLSTM [3]	0.36%	16.97%	3.87%
DCNN [28]	−7.35%	39.47%	3.13%
GRU-M [47]	28.68%	44.37%	32.53%
ResNet06 [44]	0.90%	9.35%	2.63%
Sep-CNN [14]	−0.36%	21.87%	3.03%
SC-LSTM	0.27%	9.71%	1.23%

the training dataset, ensuring the fairness and reliability of the results. For clarity, in Table IV, optimal accuracy results and minimum model parameters are highlighted in bold font, while suboptimal accuracy results and model parameters are underlined.

The SC-LSTM method, without leveraging the gossip strategy-based optimization algorithm, exhibits enhanced diagnostic accuracy compared to alternative methods, with improvements ranging from $0.37\% - 51.73\%$ (best accuracy), $6.44\% - 81.12\%$ (worst accuracy), and $3.85\% - 55.52\%$ (average accuracy). When employing the gossip strategy-based optimization algorithm, the proposed method maintains its advantages over other methods, with improvements spanning from $0 - 23.32\%$ (best accuracy), $5.71\% - 24.77\%$ (worst accuracy), and $1.21\% - 24.22\%$ (average accuracy). The experimental findings unequivocally establish the superiority of the proposed method over the other six recently published methods in terms of fault diagnosis performance. Furthermore, as indicated in Table V, the integration of the gossip strategy-based optimization algorithm leads to overall performance enhancements across different methods, particularly notable for the worst and the average accuracy.

Concerning the model trainable parameters, it is evident that despite the competitive fault diagnosis performance of 1DCNN, DCNN, and GRU-M methods, the presence of the potential local optimum problem impedes further enhancement in the diagnostic accuracy. Conversely, the lightweight models CLSTM, ResNet06, and Sep-CNN achieve commendable results owing to their concise model structures. However, due to the structural or layer limitations inherent in these lightweight models, their diagnostic accuracy is marginally lower than that of the proposed method. In summary, the proposed method can adeptly reduce model parameters, mitigating the risk of encountering the local optimum issue while ensuring the in-depth feature extraction capability. The best, worst, and average accuracy shown in Table IV can distinctly underscore the superiority of the proposed fault diagnosis method.

To visually illustrate the efficacy of the gossip strategy-based optimization algorithm in alleviating local optimum and underscore the superiority of the proposed SC-LSTM method, t-SNE is employed to visualize the worst results of different methods in ten repeated training experiments. The visualization results presented in Fig. 6 reveal that the proposed SC-LSTM method, whether with or without the gossip strategy-based optimization algorithm, exhibits compact and well-separated data clusters, with almost all samples congregating in their respective regions. In contrast, among the other six methods without the inclusion of the gossip strategy-based optimization algorithm, 1DCNN and GRU-M evidently manifest the local optimum issue, as evidenced by numerous samples from different categories mingling in the same region. Additionally, CLSTM, DCNN, ResNet06, and Sep-CNN misclassify 3-4 states into the same category. With the integration of the gossip strategy-based optimization algorithm, although

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/TII.2024.3452207,

some methods still struggle to accurately distinguish a few samples, there is a notable improvement compared to scenarios without the inclusion of the gossip strategy-based optimization algorithm. Generally speaking, the proposed gossip strategy-based optimization algorithm can effectively enhance model convergence stability, mitigate local optimum, and contributes to an overall improvement in diagnostic accuracy.

C. Performance Comparison with Different Number of Training Samples

In high-speed train operation scenarios, acquiring sufficient fault data poses a considerable challenge, primarily due to stringent safety considerations. Our study strategically addresses this limitation by employing a random selection of varying training sample sizes, ranging from 500 to 4000, to thoroughly evaluate the fault diagnosis performance of both the proposed method and six other approaches. Additionally, we ensure that each sample is unique, thereby avoiding any potential biases caused by duplicate samples. To fortify the reliability of our findings, ten repeated experiments are still conducted to mitigate the impact of neural network randomness on diagnostic results. The average diagnostic accuracy of these methods under various training sample sizes is depicted in Fig. 7, with detailed comparisons presented in Table VI.

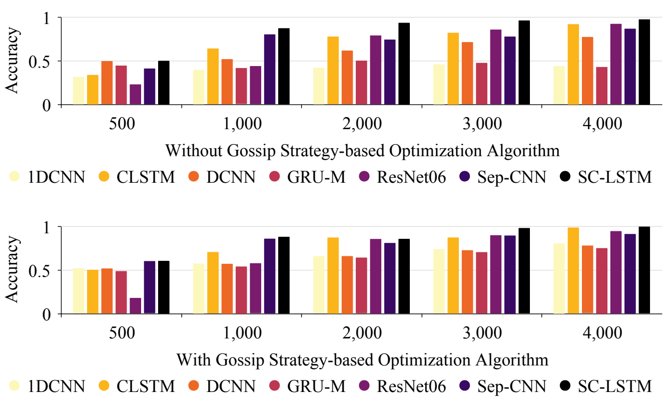

Fig. 7. Performance of different methods on HIL high-speed train traction control system simulation platform with different training samples.

TABLE VI
PERFORMANCE OF DIFFERENT METHODS ON HIL HIGH-SPEED TRAIN TRACTION CONTROL SYSTEM SIMULATION PLATFORM WITH DIFFERENT TRAINING SAMPLES.

Methods	500	1000	2000	3000	4000
Without gossip strategy-based optimization algorithm					
1DCNN [17]	31.39%	39.44%	41.86%	45.77%	43.86%
CLSTM [3]	33.61%	63.88%	77.53%	81.87%	91.71%
DCNN [28]	49.43%	51.70%	61.51%	71.18%	77.08%
GRU-M [47]	44.26%	41.53%	50.01%	47.45%	42.73%
ResNet06 [44]	22.83%	43.78%	78.80%	85.52%	92.08%
Sep-CNN [14]	40.98%	80.05%	74.02%	77.54%	86.54%
SC-LSTM	**49.79%**	**87.09%**	**93.28%**	**95.96%**	**97.22%**
With gossip strategy-based optimization algorithm					
1DCNN [17]	51.99%	57.16%	65.69%	73.76%	80.34%
CLSTM [3]	50.08%	70.39%	86.99%	87.05%	98.42%
DCNN [28]	51.60%	56.93%	65.75%	72.48%	77.80%
GRU-M [47]	48.59%	53.82%	64.03%	70.28%	74.86%
ResNet06 [44]	17.89%	57.60%	85.34%	89.61%	94.32%
Sep-CNN [14]	60.04%	85.69%	80.78%	89.27%	91.04%
SC-LSTM	**60.23%**	**87.84%**	**90.61%**	**97.86%**	**99.47%**

TABLE VII
PERFORMANCE ENHANCEMENT BY INCORPORATING GOSSIP STRATEGY-BASED OPTIMIZATION IN DIFFERENT METHODS WITH DIFFERENT TRAINING SAMPLES.

Methods	500	1000	2000	3000	4000
1DCNN [17]	20.60%	17.72%	23.83%	27.99%	36.48%
CLSTM [3]	16.74%	6.51%	9.46%	5.18%	6.71%
DCNN [28]	2.17%	5.23%	4.24%	1.30%	0.72%
GRU-M [47]	4.33%	12.29%	14.02%	22.83%	32.13%
ResNet06 [44]	−4.94%	13.82%	6.54%	4.09%	2.24%
Sep-CNN [14]	19.06%	5.64%	6.76%	11.73%	4.50%
SC-LSTM	10.44%	0.75%	−2.67%	1.90%	2.25%

It can be observed that without using the gossip strategy-based optimization algorithm, the SC-LSTM method exhibits an average diagnostic accuracy improvement of at least 0.36%, 7.04%, 14.48%, 10.44%, and 5.14% compared to the other methods for training sample scales of 500, 1000, 2000, 3000, and 4000, respectively. Upon the integration of the gossip strategy-based optimization algorithm to augment the convergence process, the SC-LSTM method continues to exhibit superiority with improvements of at least 0.19%, 2.15%, 3.62%, 8.25%, and 1.05% over other methods across different training sample scales. Notably, the proposed gossip strategy-based optimization algorithm can consistently contribute to enhancing the overall diagnostic accuracy of each model under various sample scales, as detailed in Table VII. In general, the experimental results clearly affirm that the proposed method can effectively address the local optimum issue, thereby showcasing positive performance in high-speed train fault diagnosis across diverse training sample scales.

D. Performance Comparison with Different Optimization Algorithms

To validate the superiority of the proposed gossip strategy-based optimization algorithm over other optimization algorithms in high-speed train fault diagnosis, this study introduces several well-known optimization algorithms based on the review [32], including: stochastic gradient descent with momentum (SGDM), Nesterov accelerated gradient (NAG), root mean square propagation (RMSprop), adaptive moment estimation (Adam), adaptive gradient (Adagrad), and adaptive delta (Adadelta). In the experiments conducted in this study, the model parameters are initialized using the Xavier uniform (Xa-U) initialization. Furthermore, recognizing the potential impact of different initialization methods on model convergence and local optimum, this study refers to [19] and [25] to select several well-known initialization methods for a comprehensive evaluation of the above optimization algorithms. These methods include random, LeCun normal (Le-N), LeCun uniform (Le-U), He normal (He-N), He uniform (He-U), Xavier normal (Xa-N), and Xa-U initialization.

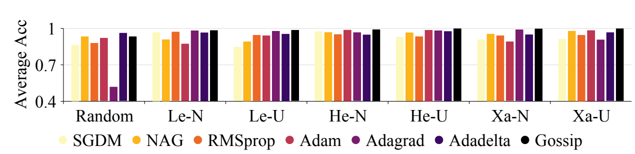

Fig. 8. Performance of different optimization algorithms on the proposed SC-LSTM model.

TABLE VIII
PERFORMANCE OF DIFFERENT OPTIMIZATION ALGORITHMS WITH DIFFERENT INITIALIZATION APPROACHES ON THE PROPOSED SC-LSTM MODEL.

Methods	Random	Le-N	Le-U	He-N	He-U	Xa-N	Xa-U
SGDM	86.15%	96.63%	84.58%	97.33%	92.75%	90.64%	90.99%
NAG	93.16%	90.69%	88.93%	96.64%	96.41%	95.28%	97.68%
RMSprop	87.75%	97.03%	94.24%	94.94%	93.19%	93.88%	94.32%
Adam	91.96%	87.17%	93.87%	98.55%	98.42%	88.83%	98.22%
Adagrad	51.68%	98.03%	97.64%	96.54%	97.93%	98.79%	90.54%
Adadelta	**96.07%**	96.37%	95.25%	94.64%	97.46%	94.78%	96.55%
Gossip	93.39%	**98.25%**	**98.35%**	**98.93%**	**99.66%**	**99.55%**	**99.66%**

Given the remarkable performance of the SC-LSTM model in high-speed train fault diagnosis, the selected optimization algorithms are deployed on the SC-LSTM model under different parameter initialization conditions to verify their performance. To minimize the impact of randomness in neural network training, each method undergoes ten repeated experiments, and the average diagnostic accuracy is presented in Fig. 8 and Table VIII. It can be seen that the proposed gossip strategy-based optimization algorithm can consistently achieve satisfactory results under different parameter initialization conditions. Particularly under Le-N, Le-U, He-N, He-U, Xa-N, and Xa-U initialization conditions, the average accuracy compared to other optimization algorithms increases by 0.22%-11.08%, 0.71%-13.77%, 0.38%-4.29%, 1.24%-6.91%, 0.76%-10.72%, and 1.44%-9.12%, respectively. Notably, Adagrad performs relatively poorly under random initialization, possibly due to the significant variation in initial gradients, making it challenging to adjust the learning rate based on past gradients. Additionally, the complex and ill-conditioned optimization landscapes may hinder Adagrad's effective convergence. The experimental results provide clear evidence that the proposed gossip strategy-based optimization algorithm can effectively alleviate local optimum and has excellent convergence ability in high-speed train fault diagnosis.

TABLE IX
DETAILED INFORMATION OF DIFFERENT OPTIMIZATION ALGORITHMS ON THE PROPOSED SC-LSTM METHOD.

Methods	Worst Acc	Training Loss	Test Loss	Epochs	Total Duration
SGDM	46.46%	0.9119	0.9475	**29**	**29.332**s
NAG	91.11%	**0.0020**	0.0980	198	199.584s
RMSprop	77.59%	0.0556	1.3831	90	110.160s
Adam	88.48%	0.0096	0.5193	188	189.504s
Adagrad	29.76%	0.6466	3.0858	132	128.304s
Adadelta	86.66%	0.0848	0.7903	68	68.544s
Gossip	**99.18%**	0.0063	**0.0141**	227	228.816s

To convincingly illustrate the effectiveness of the gossip strategy-based optimization algorithm, this study compares the proposed algorithm with several renowned optimization algorithms in terms of convergence performance, as shown in Table IX. During the model training process, we monitor the training accuracy metric. Once the training accuracy ceases to improve within a certain number of epochs, the best-performing model on the training set is applied to the test set for fault diagnosis. Table IX summarizes the detailed information of the experiment with the worst diagnostic performance on the test set under the condition of Xa-U initialization for different optimization algorithms in ten repeated training runs. The information includes the worst accuracy, final training loss, test loss, the number of epochs, and the total duration. It can be observed that the gossip strategy-based optimization algorithm achieves competitive results in terms of worst diagnostic accuracy, final training loss, and test loss. Additionally, due to the need for gossip strategy-based optimization algorithm to synthesize information saved by various recorders for further gradient descent during the training process, it requires more epochs for model convergence, resulting in longer processing time. Overall, the proposed gossip strategy-based optimization algorithm exhibits more effective convergence performance and stronger capability to avoid local optimum compared to other algorithms.

IV. CONCLUSION

To enhance the diagnostic accuracy of high-speed train traction motors, this paper proposes an innovative gossip strategy-based fault diagnosis approach, effectively mitigating the local optimum problem. The SC-LSTM introduced in this method has a lightweight model structure, and the proposed gossip strategy-based optimization algorithm can effectively prevent the local optimum problem by facilitating information sharing among multiple recorders. Several comprehensive comparison experiments are conducted, pitting the proposed gossip strategy-based SC-LSTM against six recently published methods. The performance of each method is validated on the HIL high-speed train traction control system simulation platform. The diagnostic results clearly indicate the superiority of the proposed method over the other six methods. The experimental findings also illustrate that the proposed gossip strategy-based optimization algorithm can reasonably enhance the convergence process, effectively addressing the local optimum issue and thereby improving diagnostic accuracy. For future works, it is worthwhile to apply the proposed method to similar yet different mechanical fault diagnosis tasks, such as gearbox, engine, and turbine diagnostics. Additionally, to address the time-consuming nature of the proposed method, further improvements to the gossip strategy-based optimization algorithm, such as incorporating momentum factors or adaptive modules, could accelerate model convergence and enhance diagnostic performance.

REFERENCES

[1] A. Bestavros, "Load Profiling in Distributed Real-Time Systems," Information Sciences, vol. 101, pp. 1-27, 1997.
[2] M. Cai, X. He, and D. Zhou, "Performance-Improved Finite-Time Fault-Tolerant Control for Linear Uncertain Systems with Intermittent Faults: An Overshoot Suppression Strategy," International Journal of Systems Science, vol. 53, no. 16, pp. 3408-3425, 2022.
[3] H. Chen, J. Cen, Z. Yang, W. Si, and H. Cheng, "Fault Diagnosis of the Dynamic Chemical Process Based on the Optimized CNN-LSTM Network," ACS Omega, vol. 7, no. 38, pp. 34389-34400, 2022.
[4] L. Chen, Y. Li, and S. Tong, "Neural Network Adaptive Consensus Control for Nonlinear Multi-Agent Systems Encountered Sensor Attacks," International Journal of Systems Science, vol. 54, no. 12, pp. 2536-2550, 2023.
[5] X. Chen, R. Yang, Y. Xue, M. Huang, R. Ferrero, and Z. Wang, "Deep Transfer Learning for Bearing Fault Diagnosis: A Systematic Review Since 2016," IEEE Transactions on Instrumentation and Measurement, vol. 72, 2023.
[6] W. Dong, M. Woźniak, J. Wu, W. Li, and Z. Bai, "Denoising Aggregation of Graph Neural Networks by Using Principal Component Analysis," IEEE Transactions on Industrial Informatics, vol. 19, no. 3, pp. 2385-2394, 2023.

[7] S. Feng, X. Li, S. Zhang, Z. Jian, H. Duan, and Z. Wang, "A Review: State Estimation based on Hybrid Models of Kalman Filter and Neural Network," Systems Science & Control Engineering, vol. 11, no. 1, 2023.

[8] F. Fu, D. Wang, W. Li, and F. Li, "Data-Driven Fault Identifiability Analysis for Discrete-Time Dynamic Systems," International Journal of Systems Science, vol. 51, no. 2, pp. 404-412, 2020.

[9] S. Gao, Z. Pei, Y. Zhang, and T. Li, "Bearing Fault Diagnosis Based on Adaptive Convolutional Neural Network With Nesterov Momentum," IEEE Sensors Journal, vol. 21, no. 7, pp. 9268-9276, 2021.

[10] S. Geng, C. Zhu, Y. Jin, L. Wang, and H. Tan, "Gaze control system for tracking Quasi-1D high-speed moving object in complex background," Systems Science & Control Engineering, vol. 10, no. 1, pp. 367-376, 2022.

[11] F. Han, J. Liu, J. Li, J. Song, M. Wang, and Y. Zhang, "Consensus Control for Multi-Rate Multi-Agent Systems with Fading Measurements: the Dynamic Event-Triggered Case," Systems Science & Control Engineering, vol. 11, no. 1, 2023.

[12] X. Hao, Y. Zheng, L. Lu, and H. Pan, "Research on Intelligent Fault Diagnosis of Rolling Bearing Based on Improved Deep Residual Network," Applied Sciences, vol. 11, no. 22, 2021.

[13] G. He, M. Hu, and Z. Shen, "Consensus of Switched Multi-Sgents System with Cooperative and Competitive Relationship," Systems Science & Control Engineering, vol. 11, no. 1, 2023.

[14] L. Hou, L. Liu, and G. Mao, "Machine Fault Diagnosis Method Using Lightweight 1-D Separable Convolution and WSNs With Sensor Computing," IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-8, 2022.

[15] D. Ji, C. Wang, J. Li, and H. Dong, "A Review: Data Driven-based Fault Diagnosis and RUL Prediction of Petroleum Machinery and Equipment," Systems Science & Control Engineering, vol. 9, no. 1, pp. 724-747, 2021.

[16] Y. Jiang, X. Li, C. Qin, X. Xing, and Z. Chen, "Improved Particle Swarm Optimization Based Selective Harmonic Elimination and Neutral Point Balance Control for Three-Level Inverter in Low-Voltage Ride-Through Operation," IEEE Transactions on Industrial Informatics, vol. 18, no. 1, pp. 642-652, 2022.

[17] L. Kou, Y. Qin, X. Zhao, and X. Chen, "A Multi-Dimension End-to-End CNN Model for Rotating Devices Fault Diagnosis on High-Speed Train Bogie," IEEE Transactions on Vehicular Technology, vol. 69, no. 3, pp. 2513-2524, 2020.

[18] S. Lee and T. Kim, "Impact of Deep Learning Optimizers and Hyperparameter Tuning on the Performance of Bearing Fault Diagnosis," IEEE Access, vol. 11, pp. 55046-55070, 2023.

[19] H. Li, M. Krcek, and G. Perin, "A Comparison of Weight Initializers in Deep Learning-Based Side-Channel Analysis," Applied Cryptography and Network Security Workshops, ACNS 2020, vol. 12418, pp. 126-143, 2020.

[20] X. Li, Q. Song, Z. Zhao, Y. Liu, and F. E. Alsaadi, "Optimal Control and Zero-Sum Differential Game for Hurwicz Model Considering Singular Systems with Multifactor and Uncertainty," International Journal of Systems Science, vol. 53, no. 7, pp. 1416-1435, 2021.

[21] X. Li, M. Li, P. Yan, G. Li, Y. Jiang, H. Luo, and S. Yin, "Deep Learning Attention Mechanism in Medical Image Analysis: Basics and Beyonds," International Journal of Network Dynamics and Intelligence, 2023.

[22] X. Li, Q. Song, and Y. Liu, "Optimal Control and Non-Zero-Sum Differential Game for Hurwicz Model Considering Uncertain Dynamic Systems with Multiple Input Delays," International Journal of Systems Science, vol. 54, no. 8, pp. 1676-1693, 2023.

[23] W. Liu, L. Chen, Y. Chen, and W. Zhang, "Accelerating Federated Learning via Momentum Gradient Descent," IEEE Transactions on Parallel and Distributed Systems, vol. 31, no. 8, pp. 1754-1766, 2020.

[24] M. Ma and Z. Mao, "Deep-Convolution-Based LSTM Network for Remaining Useful Life Prediction," IEEE Transactions on Industrial Informatics, vol. 17, no. 3, pp. 1658-1667, 2021.

[25] M. V. Narkhede, P. P. Bartakke, and M. S. Sutaone, "A Review on Weight Initialization Strategies for Neural Networks," Artificial Intelligence Review, vol. 55, no. 1, pp. 291-322, 2021.

[26] N. Qin, B. Wu, D. Huang, and Y. Zhang, "Stepwise Adaptive Convolutional Network for Fault Diagnosis of High-Speed Train Bogie Under Variant Running Speeds," IEEE Transactions on Industrial Informatics, vol. 18, no. 12, pp. 8389-8398, 2022.

[27] F. M. Shakiba, M. Shojaee, S. M. Azizi, and M. Zhou, "Real-Time Sensing and Fault Diagnosis for Transmission Lines," International Journal of Network Dynamics and Intelligence, pp. 36-47, 2022.

[28] S. Shao, R. Yan, Y. Lu, P. Wang, and R. X. Gao, "DCNN-Based Multi-Signal Induction Motor Fault Diagnosis," IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 6, pp. 2658-2669, 2020.

[29] F. Song, Y. Li, W. Cheng, and L. Dong, "An Improved Dynamic Programming Tracking-Before-Detection Algorithm based on LSTM Network Value Function," Systems Science & Control Engineering, vol. 11, no. 1, 2023.

[30] W. Song, J. He, J. Lin, H. Ye, X. Ling, and C. Lu, "Bias Analysis of PMU-Based State Estimation and Its Linear Bayesian Improvement," IEEE Transactions on Industrial Informatics, vol. 20, no. 2, pp. 1607-1617, 2024.

[31] T. Sun, L. Qiao, Q. Liao, and D. Li, "Novel Convergence Results of Adaptive Stochastic Gradient Descents," IEEE Transactions on Image Processing, vol. 30, pp. 1044-1056, 2021.

[32] Y. Tian, Y. Zhang, and H. Zhang, "Recent Advances in Stochastic Gradient Descent in Deep Learning," Mathematics, vol. 11, no. 3, 2023.

[33] C. Wang, Z. Wang, W. Liu, Y. Shen, and H. Dong, "A Novel Deep Offline-to-Online Transfer Learning Framework for Pipeline Leakage Detection With Small Samples," IEEE Transactions on Instrumentation and Measurement, vol. 72, pp. 1-13, 2023.

[34] C. Wang, Z. Wang, H. Liu, H. Dong, and G. Lu, "An Optimal Unsupervised Domain Adaptation Approach With Applications to Pipeline Fault Diagnosis: Balancing Invariance and Variance," IEEE Transactions on Industrial Informatics, pp. 1-12, 2024.

[35] J. Wang, Y. Zhuang, and Y. Liu, "FSS-Net: A Fast Search Structure for 3D Point Clouds in Deep Learning," International Journal of Network Dynamics and Intelligence, 2023.

[36] T. Wang, Z. Zhang, and K.-L. Tsui, "A Deep Generative Approach for Rail Foreign Object Detections via Semisupervised Learning," IEEE Transactions on Industrial Informatics, vol. 19, no. 1, pp. 459-468, 2023.

[37] W. Wang, L. Xu, J. Xu, X. Li, and H. Zhang, "Linear Quadratic Optimal Control for Time-Delay Stochastic System with Partial Information," International Journal of Systems Science, vol. 54, no. 10, pp. 2227-2238, 2023.

[38] Z. Wang, S. Tian, H. Gao, C. Han, and F. Guo, "An On-Line Detection Method and Device of Series Arc Fault Based on Lightweight CNN," IEEE Transactions on Industrial Informatics, vol. 19, no. 10, pp. 9991-10003, 2023.

[39] J. Xu, H. Ke, Z. Chen, X. Fan, T. Peng, and C. Yang, "Oversmoothing Relief Graph Convolutional Network-Based Fault Diagnosis Method With Application to the Rectifier of High-Speed Trains," IEEE Transactions on Industrial Informatics, vol. 19, no. 1, pp. 771-779, 2023.

[40] Y. Xue, R. Yang, X. Chen, Z. Tian, and Z. Wang, "A Novel Local Binary Temporal Convolutional Neural Network for Bearing Fault Diagnosis," IEEE Transactions on Instrumentation and Measurement, , vol. 72, 2023.

[41] C. Yang, C. Yang, T. Peng, X. Yang, and W. Gui, "A Fault-Injection Strategy for Traction Drive Control Systems," IEEE Transactions on Industrial Electronics, vol. 64, no. 7, pp. 5719-5727, 2017.

[42] J. Yang, S. Yin, Y. Chang, and T. Gao, "A Fault Diagnosis Method of Rotating Machinery Based on One-Dimensional, Self-Normalizing Convolutional Neural Networks," Sensors (Basel), vol. 20, no. 14, 2020.

[43] X. Yi, H. Yu, Z. Fang, and L. Ma, "Probability-Guaranteed State Estimation for Nonlinear Delayed Systems Under Mixed Attacks," International Journal of Systems Science, vol. 54, no. 9, pp. 2059-2071, 2023.

[44] S. Yu, M. Wang, S. Pang, L. Song, and S. Qiao, "Intelligent Fault Diagnosis and Visual Interpretability of Rotating Machinery based on Residual Neural Network," Measurement, vol. 196, 2022.

[45] X. Zhai, F. Qiao, Y. Ma, and H. Lu, "A Novel Fault Diagnosis Method Under Dynamic Working Conditions Based on a CNN With an Adaptive Learning Rate," IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-12, 2022.

[46] J. Zhang, G. Wan, M. Jiang, G. Lu, X. Tao, and Z. Huang, "Small Object Detection in UAV Image based on Improved YOLOv5," Systems Science & Control Engineering, vol. 11, no. 1, 2023.

[47] Y. Zhang, T. Zhou, X. Huang, L. Cao, and Q. Zhou, "Fault Diagnosis of Rotating Machinery based on Recurrent Neural Networks," Measurement, vol. 171, 2021.

[48] M. Zhong, X. Zhu, T. Xue, and L. Zhang, "An Overview of Recent Advances in Model-based Event-Triggered Fault Detection and Estimation," International Journal of Systems Science, vol. 54, no. 4, pp. 929-943, 2022.