Journal of Dynamics, Monitoring and Diagnostics, 2025, 4, 203-212
https://doi.org/10.37965/jdmd.2025.871

ISTP
RESEARCH ARTICLE

Interpretable Fault Diagnosis for Liquid Rocket Engines via Component-Wise MLP-Based Granger Causality Feature Extraction

Longfei Zhang,[1,2] Zhi Zhai,[1,2] Chenxi Wang,[1,2] Meng Ma,[1,2] Jinxin Liu,[1,2] and Chunmin Wang[1,2,3]

[1]School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, P.R. China
[2]National Key Lab of Aerospace Power System and Plasma Technology, Xi'an Jiaotong University, Xi'an 710049, P.R. China
[3]Xi'an Aerospace Propulsion Institute, Xi'an 710100, P.R. China

(Received 15 July 2025; Revised 15 August 2025; Accepted 14 September 2025; Published online 16 September 2025)

Abstract: Liquid rocket engine (LRE) fault diagnosis is critical for successful space launch missions, enabling timely avoidance of safety hazards, while accurate post-failure analysis prevents subsequent economic losses. However, the complexity of LRE systems and the "black-box" nature of current deep learning-based diagnostic methods hinder interpretable fault diagnosis. This paper establishes Granger causality (GC) extraction-based component-wise multi-layer perceptron (GCMLP), achieving high fault diagnosis accuracy while leveraging GC to enhance diagnostic interpretability. First, component-wise MLP networks are constructed for distinct LRE variables to extract inter-variable GC relationships. Second, dedicated predictors are designed for each variable, leveraging historical data and GC relationships to forecast future states, thereby ensuring GC reliability. Finally, the extracted GC features are utilized for fault classification, guaranteeing feature discriminability and diagnosis accuracy. This study simulates six critical fault modes in LRE using Simulink. Based on the generated simulation data, GCMLP demonstrates superior fault localization accuracy compared to benchmark methods, validating its efficacy and robustness.

Keywords: fault diagnosis; Granger causality; interpretability; liquid rocket engine; MLP

I. INTRODUCTION

In recent years, major spacefaring nations have proposed significant aerospace initiatives, including satellite constellations [1], deep space exploration, and manned lunar landings [2]. Liquid rocket engines (LREs) are positioned as strategic priorities for current and future aerospace systems [3], with goals centered on high reliability, low cost, and reusability. Characterized by structural complexity, extreme operating conditions, and intricate dynamic behaviors, LREs constitute a failure-prone component in launch vehicles. Any malfunction could induce catastrophic impacts on space missions [4]. Consequently, comprehensive research on LRE fault diagnosis is imperative to enhance operational safety and ensure mission success.

Currently, prevalent fault diagnosis methods for LREs primarily fall into three categories: model-driven, data-driven, and knowledge-driven approaches [5]. The model-driven approaches establish mathematical or physical models of the engine system based on its operational principles under normal working conditions. Fault identification is achieved by analyzing residuals between the model's output values and the engine's actual measurements under identical input conditions. Kawatsu *et al.* [6] developed a fault detection method for LRE electromechanical actuators based on multi-physics system-level modeling and simulation combined with the Dynamic Time Warping (DTW) algorithm. Cha *et al.* [7] proposed a fault detection and diagnosis algorithm based on nonlinear Kalman filtering for open-cycle LREs, enabling reliable transient-state fault localization. Xu *et al.* [8] employed the Unscented Kalman Filter (UKF) to achieve accurate identification of three types of faults in a LRE. Sun *et al.* [9] proposed a multiple-model-based fault sensor isolation method for LRE that integrated model identification techniques and particle filter bank, enabling effective isolation of open-circuit faults and drift faults in sensors.

Knowledge-based fault diagnosis methods rely on fault patterns, failure modes, and expert experience summarized from historical anomaly cases to determine the operating state of the engine. Representative of this category are expert systems, which are typically composed of a knowledge base, an inference engine, and an interpreter. The knowledge base is responsible for storing knowledge, the inference engine derives conclusions based on the stored knowledge, and the interpreter explains the system's behavior to users. In the context of LRE, expert systems applied in engineering practice include Aerojet Propulsion's Titan system, SPARTA embedded expert system, and Rocketdyne's turbopump fault diagnosis system. The primary advantage of such methods lies in their strong interpretability and independence from precise mathematical models; however, their main limitation is the difficulty of acquiring expert knowledge and maintaining rule sets.

Data-driven approaches bypass the need to comprehend the inherent complexity of engines, instead,

Corresponding author: Chenxi Wang (e-mail: wangchenxi@xjtu.edu.cn)

identifying faults and locating their sources by analyzing correlations between engine measurement signals and fault manifestations. For LREs with intricate structures and operational processes, where establishing concise and accurate system models is often infeasible, data-driven methods remain the predominant approach in the field of LRE fault detection and diagnosis. Sun *et al.* [10] achieved real-time fault detection for LREs by determining fault-characteristic frequency bands through spectral analysis and using the Root-Mean-Square (RMS) values of these bands as detection indicators. Deng *et al.* [11], in their study on main-stage fault diagnosis for a high-thrust hydrogen–oxygen staged-combustion cycle engine, designed a real-time fault detection method based on an autoregressive moving average (ARMA) model for the steady-state phase and verified its reliability via a hardware-in-the-loop simulation platform.

Recent breakthroughs in big data and artificial intelligence (AI) have enabled deep learning-based intelligent fault diagnosis methods to leverage hidden information in data, which demonstrates significant advantages across domains. Chen *et al.* [12] developed a physics-informed deep neural network for bearing remaining useful life prediction. Wu *et al.* [13] proposed CA-DenseNet for fault diagnosis in linear guideways. Such methods have also gained significant traction in LRE diagnostics. Park *et al.* [14] proposed a fault detection and diagnosis method for liquid-propellant rocket engine tests during startup transient based on a convolutional neural network-long short-term memory (CNN-LSTM). Wang *et al.* [15] addressed the scarcity of fault patterns in LREs during actual operation by leveraging a pre-training and fine-tuning framework based on a CNN. Yan *et al.* [16] combined a bidirectional RNN and attention assistance based on multi-granularity reference to achieve fault localization in the launch vehicle test and launch process.

Analysis of LRE systems and associated fault diagnosis methodologies reveals three fundamental challenges: (1) LRE sensor data exhibit three defining characteristics: temporal dependencies reflecting sequential patterns in sensor evolution, high dimensionality arising from complex multivariate measurements across numerous sensors, and nonlinear interactions where interdependencies violate linear assumptions. These characteristics collectively impede high-precision fault diagnosis in LRE. (2) Model-driven diagnostic methods for LREs rely on accurate physical or mathematical models, which can provide physically meaningful information and effectively detect unknown fault modes. However, these methods demand a model with high precision, a significant challenge due to the difficulty in constructing high-fidelity LRE models. Moreover, variations between individual engines necessitate separate modeling or adjustments for each engine to ensure accuracy, resulting in poor generalizability and robustness of models. (3) In recent years, AI-powered data-driven diagnostic methods demonstrate powerful feature learning capabilities, strong nonlinear fitting capacity, and adaptability independent of specific systems or devices. Yet, as neural networks inherently function as black-box models, they fail to incorporate human-interpretable semantics into the feature space, leading to limited interpretability of diagnostic results. This poses a potential risk for LREs, where high reliability is paramount.

To address the aforementioned challenges, Granger causality (GC) [17]-based causal learning provides a solution. GC defines causality from a predictive perspective: if variable a contributes statistically significant improvement in predicting variable b, then a is regarded Granger causal for b. GC inherently depends on the activity of the entire system of time series under study, making it more appropriate for understanding high-dimensional complex data streams [18], and thus applicable to fault diagnosis in LRE. Constructing GC models among time series could circumvent the stringent requirement of model-driven methods for high-precision mathematical or physical models of LRE. Simultaneously, GC structure inherently characterizes directional dependencies between temporal variables, thereby eliminating the opacity of conventional deep learning methods and establishing explainable diagnostic reasoning. Ma *et al.* [19] proposed attention-based random disturbance gated recurrent unit (ARDGRU) for nonlinear dynamic GC analysis, achieving root cause analysis in manufacturing processes. Han *et al.* [20] proposed a novel autoencoder-based framework for root cause analysis (AERCA) leveraging GC discovery to achieve root cause analysis in multivariate time series anomalies. Zhang *et al.* [21] proposed a graph neural network method based on the GC test for bearing fault detection, achieving accurate fault classification. The application of GC-based deep learning in diverse fields provides methodological insights for addressing the aforementioned challenges in LRE fault diagnostics.

Therefore, this paper proposes a GC-based deep learning method for fault classification in LREs. Since distinct faults correspond to unique propagation pathways [22], manifested as divergent GC relationships, we extract GC as discriminative features for fault classification. To address the strong coupling among LRE variables, a component-wise strategy is adopted; therefore, GC effects on individual variables are separately extracted and integrated to form a complete GC model. Combined with the above analysis, considering that many scholars have utilized MLP to extract GC and achieved good results in subsequent tasks [23–25], this paper employs MLP as the backbone of the network and proposes GC extraction-based component-wise multi-layer perceptron (GCMLP) for LRE fault diagnosis. By utilizing data influenced by GC for prediction, enhanced forecasting accuracy validates the rationality and correctness of extracted GC features; employing GC characteristics for fault classification ensures discriminability across different faults through improved diagnostic accuracy, while imposing sparsity constraints on the extracted GC structure aligns with the intrinsic sparse connectivity characteristics of causal models. The primary contributions of this work are summarized as follows:

(1) Methodological innovation: we propose a novel GC-MLP framework based on an encoder–decoder structure, enabling both accurate classification and interpretable feature extraction.

(2) Causality-enhanced diagnosis: we introduce GC-based features to capture failure propagation mechanisms in LREs, bridging data-driven learning with physical interpretability.

(3) Comprehensive validation: we demonstrate the effectiveness on LRE simulation data, compare with multiple baseline models (GRU, BiGRU, LSTM, BiLSTM, CNN, and MLP), and analyze the effect of different sparsity constraints on diagnostic accuracy.

II. PRELIMINARY KNOWLEDGE

This section will introduce GC and explain how sparsity constraints are imposed on the GC matrix during the extraction process, which serves as the foundational framework for the proposed method.

A. GRANGER CAUSALITY

Granger defined causality based on whether past values of a time series x_t contribute to predicting future values of another time series y_t. Let $H_{<t}$ denote all relevant information available up to time $t-1$, and $P(y_t|H_{<t})$ represent the prediction of y_t on the basis of $H_{<t}$. If

$$\mathrm{var}[y_t - P(y_t|H_{<t})] < \mathrm{var}[y_t - P(y_t|H_{<t}\backslash\, x_{<t})] \qquad (1)$$

x_t is Granger causal for y_t, where $H_{<t}\backslash\, x_{<t}$ denotes all information in $H_{<t}$ excluding the past values of the time series x_t.

Early GC was defined based on linear relationships among variables. Given a time series vector $x = (x_1, x_2, \ldots, x_T)$ in dimensions of (p,T), where $x_t = (x_t^1, x_t^2, \ldots, x_t^p)^T$ where p denotes the number of sensors, T denotes the length of time series, and t denotes the t-th time point. The components satisfy the following linear relationship:

$$A^0 x_t = \sum_{k=1}^{d} A^k x_{t-k} + e_t \qquad (2)$$

where $A^0, A^1, \ldots, A^d$ is a causal coefficient matrix of dimensions $p \times p$, representing GC relationships. $A_{ij}^k \neq 0$ indicates that time series x^j is Granger causal for x^i, d denotes lag, and e_t denotes noise. Taking the GC graph among the four time series variables shown in Fig. 1 as an example, the GC effects vary across different time lags where lag=2. The corresponding causal adjacency matrix for a specific time lag is shown in Fig. 2. This formulation can also display nonlinear GC relationships.

B. LASSO

For high-dimensional time series data, the causal structure among variables typically exhibits sparsity. When employing neural networks to capture GC between multiple time series variables, a sparsity-based regularization term is commonly introduced into the loss function. This ensures sparse connectivity between variables, thereby obtaining a reasonable causality estimation. Fujita *et al.* [26] introduced lasso penalty to enforce sparsity in the GC matrix, driving most elements to zero. The formulation is given by:

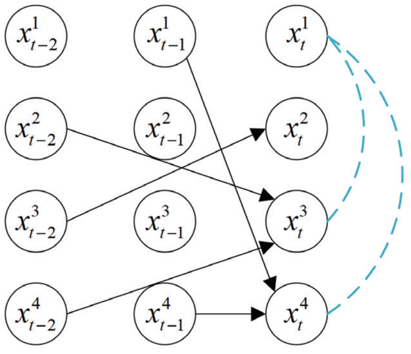

Fig. 1. GC diagram.

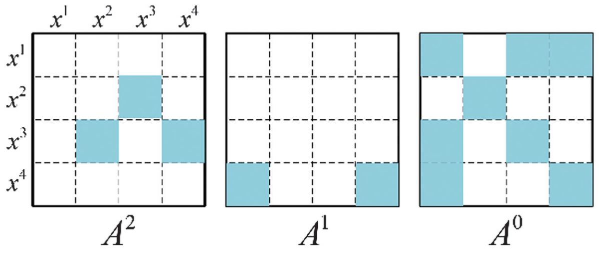

Fig. 2. GC matrix.

$$\Omega(A^1, A^2, \ldots, A^d) = \lambda \sum_{k=1}^{d} \sum_{i=1}^{i} \sum_{j=1}^{p} |A_{ij}^k| \qquad (3)$$

where Ω is a penalty term restricting the sparsity of GC. Shojaie *et al.* [27] proposed the truncating lasso penalty, which computes weights based on causal matrices from prior lags to adjust the sparsity penalty for subsequent lag matrices, The penalty is defined as below, where Ψ^k is the weight corresponding to GC at lag k, M is a large constant, and β is the user-specified tolerable false negative rate:

$$\Omega(A^1, A^2, \ldots, A^d) = \lambda \sum_{k=1}^{d} \Psi^k \sum_{i=1}^{i} \sum_{j=1}^{p} |A_{ij}^k| \qquad (4)$$

$$\Psi^1 = 1, \Psi^k = M^{I\{\|A^{k-1}\|_0 < p^2\beta/(d-k)\}}, \;\; k \geq 2 \qquad (5)$$

Lazano *et al.* [28] introduced group lasso penalty:

$$\Omega(A^1, A^2, \ldots, A^d) = \lambda \sum_{i=1}^{p} \sum_{j=1}^{p} |(A_{ij}^1, A_{ij}^2, \ldots, A_{ij}^d)|_2 \qquad (6)$$

Basu *et al.* [29] proposed a more generalized group lasso penalty incorporating relationships between variables. Unlike grouping across lags, their approach treats arbitrary subsets of variables or entire matrices as a single group for L2 norm computation. The formulation is expressed as:

$$\Omega(A^1, \ldots, A^d) = \lambda \sum_{k=1}^{d} \sum_{i=1}^{p} |(A_{i1}^k, A_{i2}^k, \ldots, A_{ip}^k)|_2 \qquad (7)$$

Nicholson *et al.* [30] proposed hierarchical group lasso penalty based on decay assumption. This approach enforces that if all elements in the causal matrix at lag k are zero, then elements in causal matrices for larger lag are also entirely zeros, formally:

$$A^k = 0 \Rightarrow A^m = 0, \;\; \forall m > k \qquad (8)$$

III. DESCRIPTION OF THE METHOD

A. MODEL ARCHITECTURE

The proposed network employs an integrated Encoder–Decoder architecture to extract GC of LRE variables and achieve fault classification. Therefore, the core focus of this network lies in GC acquisition, with the objective of achieving accurate classification across diverse fault types.

The framework consists of three core modules: the Encoder, serving as the GC Extracting Module, the Decoder, acting as the Predictor, and the Fault Classifier. The network architecture is illustrated in Fig. 3.

(1) The GC Extracting Module (Encoder) is designed to capture GC between variables in LRE multivariate time series. Adopting a component-wise approach, a separate MLP is designed for each variable to avoid the influence of inter-variable coupling on GC extraction. The resulting GC reflects the causal relationships between various variables in the LRE and serves as input features for subsequent fault localization.

(2) The role of Predictor (Decoder) is to predict the next-time-step values of each variable based on LRE time series data influenced by the GC matrix. During backpropagation, it provides prediction errors to assist the GC Extracting Module in parameter updating, ensuring the correctness of the extracted GC matrix so that it can accurately reflect inter-variable interactions.

(3) The Classifier performs fault localization using the GC extracted by the GC Extracting Module as input features and serves as the final output component of the fault diagnosis task. During backpropagation, it provides classification errors, which not only enhances the Classifier's ability to distinguish between different faults but also assists in optimizing the GC Extracting Module, prompting it to extract GC matrices with discriminability for different fault data. By improving the separability of input features, the accuracy of Classifier's classification could be improved.

The GC Extracting Module employs an intricately tailored component-wise architecture built upon MLP foundations. The module is designed to distill GC relationships across multivariate time-lagged data (with maximum lag d) and to encode these relationships into GC matrices systematically. And the obtained matrices serve as both a structured input for the downstream Predictor and the discriminative input features for the Fault Classifier.

Preprocessed data entering the GC Extracting Module is $X \in R^{n_b * p * d}$, where n_b denotes the count of samples in a batch, p denotes the number of sensors, and d is the maximum time lag. The component-wise paradigm manifests through p specialized GC extractor groups. Each group focuses on analyzing GC effects from all variables (including autoregressive effects) on a single-target LRE variable—generating one row of the GC matrix. Concatenation of these row vectors yields the complete GC matrix for a specific time lag. To account for temporal variations in GC across different time lags, each GC extractor group incorporates d specialized extractors that separately quantify GC effects received by a target variable at each lag step where lag $= 1,2, \ldots ,d$. The module comprises $d \times p$ parallel GC extractors, each designed to capture the GC effects of p variables across d time lags. The output is a tensor with dimensions $d \times p \times p$, where each slice along the lag dimension corresponds to a GC matrix at a specific time lag.

B. OPTIMIZATION OBJECTIVE

Since the GC between LRE sensors is established on whether lagged data from each channel contributes to predicting future values, the training process requires quantifying the discrepancy between the Predictor's outputs and the ground-truth sensor measurements. This discrepancy should progressively decrease during training to ensure the

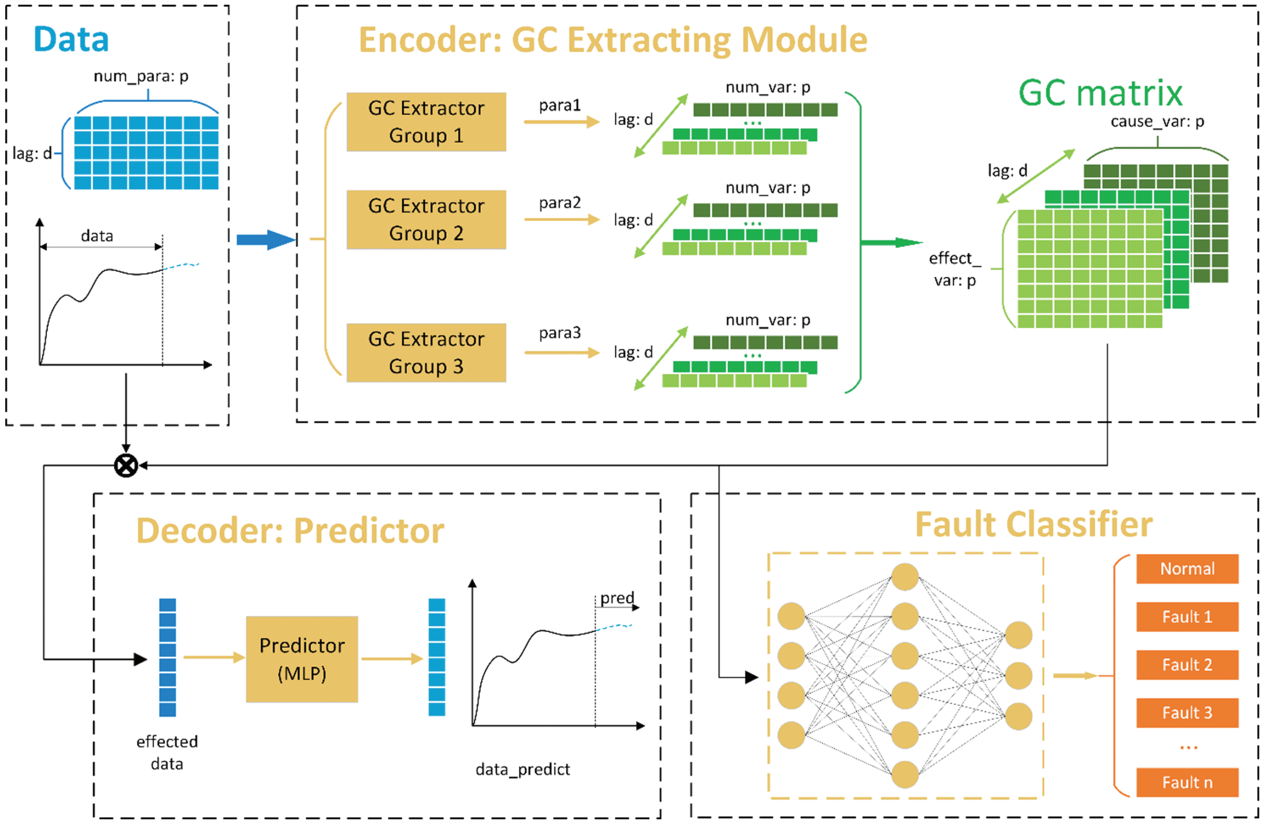

Fig. 3. Description of the method.

accuracy of the GC matrix. The mean squared error (MSE) is a widely adopted metric for quantifying prediction accuracy and is defined as:

$$L_{\text{pred}} = \frac{1}{p}\sum_{i=1}^{p}(\tilde{x}_t^i - x_t^i)^2 \tag{9}$$

where p denotes the number of sensors, x_t^i denotes measured value of the i-th sensor at time point t, and $\tilde{x}_t^i$ denotes predicted value of the i-th sensor at time point t generated by Predictor.

In LREs, sensor monitoring points are typically numerous, yet for any individual sensor, only a few other sensors exhibit GC relationships with it. To ensure sparsity of GC matrix in high-dimensional data, the group lasso penalty is introduced to constrain the GC matrix output by the GC Extractor. This regularization is formally defined as:

$$L_{\text{spar}} = \frac{1}{p}\sum_{i=1}^{p}\|A_{i:}\| = \frac{1}{p}\sum_{i=1}^{p}\sqrt{\sum_{j=1}^{p}|A_{ij}|^2} \tag{10}$$

where p denotes the number of sensors and $A_{i:}$ denotes the elements in the i-th row of the GC matrix, physically indicating whether other variables are Granger causal for the i-th variable.

Additionally, as the model's core task involves fault classification, the classification accuracy of the Fault Classifier serves as a critical performance metric. To optimize this component, a cross-entropy loss function is applied during training. This approach simultaneously enhances the classifier's diagnostic precision and boosts the discriminability of GC matrices extracted by the GC Extractor across diverse fault types ensuring distinct GC matrix patterns emerge for different failure modes. The classification loss is defined as:

$$L_{\text{class}} = -y\text{log}\tilde{y} - [(1-y)\ \log(1-\tilde{y})] \tag{11}$$

Synthesizing above considerations, the training process must simultaneously balance GC matrix accuracy, topological sparsity, inter-fault distinguishability, and fault classification precision. This yields a consolidated loss function integrating all objectives:

$$\begin{aligned} L &= \lambda_1 L_{\text{pred}} + \lambda_2 L_{\text{spar}} + \lambda_3 L_{\text{class}} \\ &= \lambda_1 * \frac{1}{p}\sum_{i=1}^{p}(\tilde{x}_t^i - x_t^i)^2 + \lambda_2 \\ &\quad * \frac{1}{p}\sum_{i=1}^{p}\sqrt{\sum_{j=1}^{p}|A_{ij}|1^2} \\ &\quad + \lambda_3 * \{-y\ \log\tilde{y} - [(1-y)\ \log(1-\tilde{y})]\} \end{aligned} \tag{12}$$

where λ_1, λ_2, and λ_3 denote the regularization weight for L_{pred} (prediction accuracy constraint), L_{spar} (sparsity loss), and L_{class} (classification penalty), respectively.

IV. EXPERIMENT VERIFICATION

A. DATA DESCRIPTION

The experimental data are obtained from an LRE simulation model of a specific type, acquiring six distinct fault datasets through fault injection techniques. This LRE model comprises multiple hierarchical subsystems, including an oxidizer supply system, valve control system, turbopump assembly system, and so on, with a schematic diagram illustrating its operational principles provided in Fig. 4.

The LRE of this specific model is decomposed into core components, including pumps, turbines, combustion chambers, and pipelines, where the strong dynamical coupling between numerous subsystems necessitates strategic simplification. A modular modeling methodology is adopted. First, physics-based dynamic models is constructed for each critical component in MATLAB/Simulink according to their underlying operational principles. Subsequently, these subsystems are integrated through systematic coupling and debugging of input–output relationships to derive the integrated system-level LRE model, as illustrated in Fig. 5.

Based on the aforementioned Simulink model, the experimental data is constructed through the following steps:

(1) Fault simulation: Faults are simulated using fault injection techniques [31] introducing specific failure modes artificially into the system during operation.

(2) Data normalization: Sensor data undergo min–max normalization to eliminate dimensional discrepancies and prevent skewed feature extraction during GC construction. The formula is as follows:

$$x^i = \frac{x^i - \min(x^i)}{\max(x^i) - \min(x^i)},\ i = 1,2\ldots p \tag{13}$$

(3) To emulate real-world environmental interference and enhance model robustness, white Gaussian noise with a 20 dB SNR is injected into normalized data.

Following data preprocessing, a dataset encompassing six distinct fault types is established, involving main turbine fault, oxygen pump fault, stage-1 fuel pump fault, thrust chamber throat ablation, oxygen booster pump fault, and fuel booster pump fault. Each fault was simulated at two severity levels, resulting in 12 distinct datasets. Every dataset contains 10 sensor channels sampled at 1 kHz over a 5-second duration, with specific sensor configurations detailed in Table I.

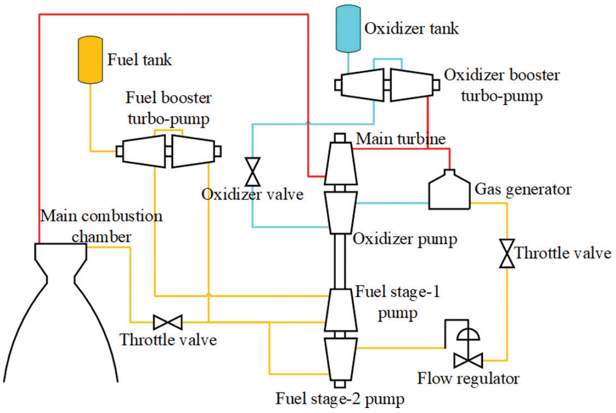

Fig. 4. Schematic diagram of a specific LRE model.

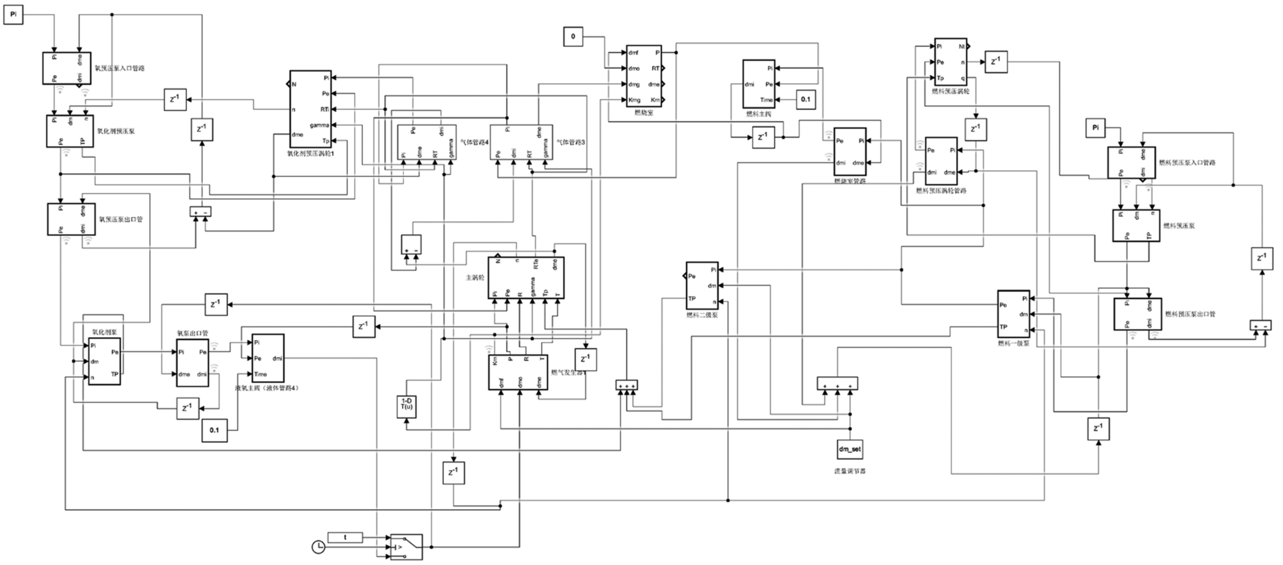

Fig. 5. Simulink model of a specific LRE type.

Table I. Sensor ID and measured parameters

No.	Sensor
1	Oxidizer pump inlet pressure
2	Oxidizer pump outlet pressure
3	Oxidizer pump flow rate
4	Fuel stage1 pump inlet pressure
5	Fuel stage1 pump outlet pressure
6	Fuel stage1 pump flow rate
7	Fuel stage2 pump outlet pressure
8	Main turbo rotational speed
9	Main turbo flow rate
10	Gas generator pressure

The operational timeline consists of three phases: normal LRE operation at 0–2.5 seconds, a precise 200-millisecond fault injection window at 2.5–2.7 seconds, and sustained fault-state operation at 2.7–5 seconds. Preprocessed sensor data for the main turbine efficiency degradation fault is partially illustrated in Fig. 6.

During dataset construction, data from the 0.5–1s interval are selected as the normal state samples. For fault data selection, the post-injection transition period (2.5–3s) is adopted. This interval captures dynamic characteristics of the LRE model during fault initiation and stabilization, facilitating effective GC extraction by GCMLP. Consequently, the 2.5–3s data represent the fault state. The raw data were segmented using sliding windows with a window length of 10 and step size of 5, yielding a final dataset comprising 1,274 samples.

B. EXPERIMENT RESULT

Table II presents the fault classification performance of various methods on the established dataset, with all accuracy values averaged over five independent experimental trials to ensure statistical stability. The proposed GCMLP method is benchmarked against six widely adopted approaches: CNN, MLP, and time series specialized architectures including gated recurrent unit (GRU), bidirectional gated recurrent unit (BiGRU), LSTM, and bidirectional long short-term memory (BiLSTM). GCMLP achieves superior metrics of 97.67% accuracy, 96.70% precision, and 96.39% recall—collectively outperforming all comparative methods and demonstrating enhanced

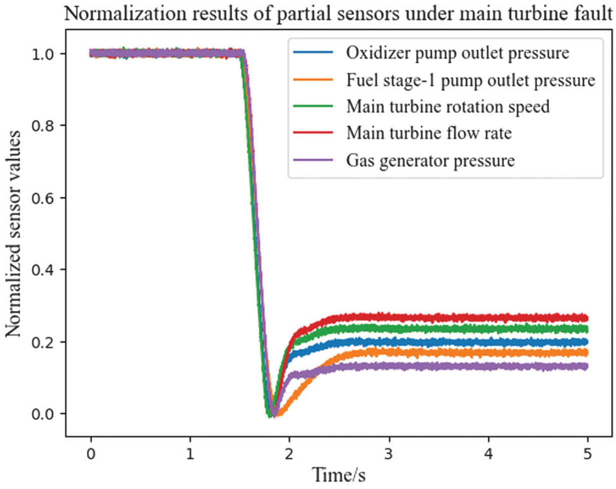

Fig. 6. Normalization results of partial sensor data under main turbine fault.

Table II. Fault classification accuracy of baseline models and GCMLP

Model	Accuracy	Precision	Recall
CNN	92.25%	88.28%	88.87%
GRU	88.22%	82.49%	86.50%
BiGRU	83.88%	74.17%	83.70%
LSTM	86.05%	78.32%	82.61%
BiLSTM	92.56%	85.83%	88.69%
MLP	89.77%	84.35%	87.47%
GCMLP	**95.50%**	**93.03%**	**94.71%**

TABLE III. Fault types and corresponding labels in confusion matrices

Label	Fault type
0	Normal
1	Main turbine fault
2	Oxidizer pump fault
3	Fuel stage-1 pump fault
4	Thrust chamber throat ablation
5	Oxidizer booster pump fault
6	Fuel booster pump fault

capability in precise fault identification and false alarm mitigation.

The confusion matrices (label indices mapped to fault types in Table III) in Fig. 7 reveal that GCMLP and five comparative methods all maintain over 90% recognition accuracy for Faults 2–6. Notably, GCMLP demonstrates exceptional performance in identifying both healthy states and fuel booster pump fault (Label 6), whereas the benchmark methods exhibit significant deficiencies in diagnosing these two categories: their accuracy for healthy states hovers around 50%, with BiLSTM achieving only 12% accuracy for fuel booster pump fault (effectively failing to detect it), while LSTM and MLP reach approximately 50%, and CNN/GRU attain 84%. In stark contrast, GCMLP achieves 100% accuracy for fuel booster pump fault. Collectively, the recognition rates across all fault types and the holistic analysis of the confusion matrix conclusively validate the efficacy of the proposed method.

The innovation of this paper lies in incorporating GC extraction, which enables causal interpretation for feature extraction and fault localization while maintaining classification accuracy. Taking thrust chamber throat ablation faults as a case study, we analyze the rationality of causal features extracted by GCMLP and the interpretability of fault diagnosis results. Figure 8 shows GC relationships among sensors during throat ablation faults in the dataset mentioned above. Figure 9 visualizes GC relationships derived from the GC matrix extracted by GCMLP. Subsequent analysis compares these relationships sequentially from left to right.

Derived from physical connections and equilibrium relationships among engine variables:

$$P_{\text{com}} \approx P_{\text{tur}} \tag{14}$$

$$P_{\text{fp1e}}= P_{\text{fp2i}} \tag{15}$$

where P_{com} denotes combustion chamber pressure, P_{tur} denotes main turbine outlet pressure, P_{fp1e} denotes fuel stage-1 pump outlet pressure, and P_{fp2i} denotes fuel stage-2 pump inlet pressure.

(1) The fuel stage-1 pump is connected to the combustion chamber via a valved pipeline, establishing topological linkage. Here, P_{com} serves as an approximation for P_{com}. Thrust chamber throat ablation manifests through three correlated effects: throat area enlargement reduces combustion chamber pressure P_{com} and simultaneously decreases main turbine outlet pressure P_{tur} due to parametric equilibrium; the consequent turbine pressure ratio elevation increases rotational

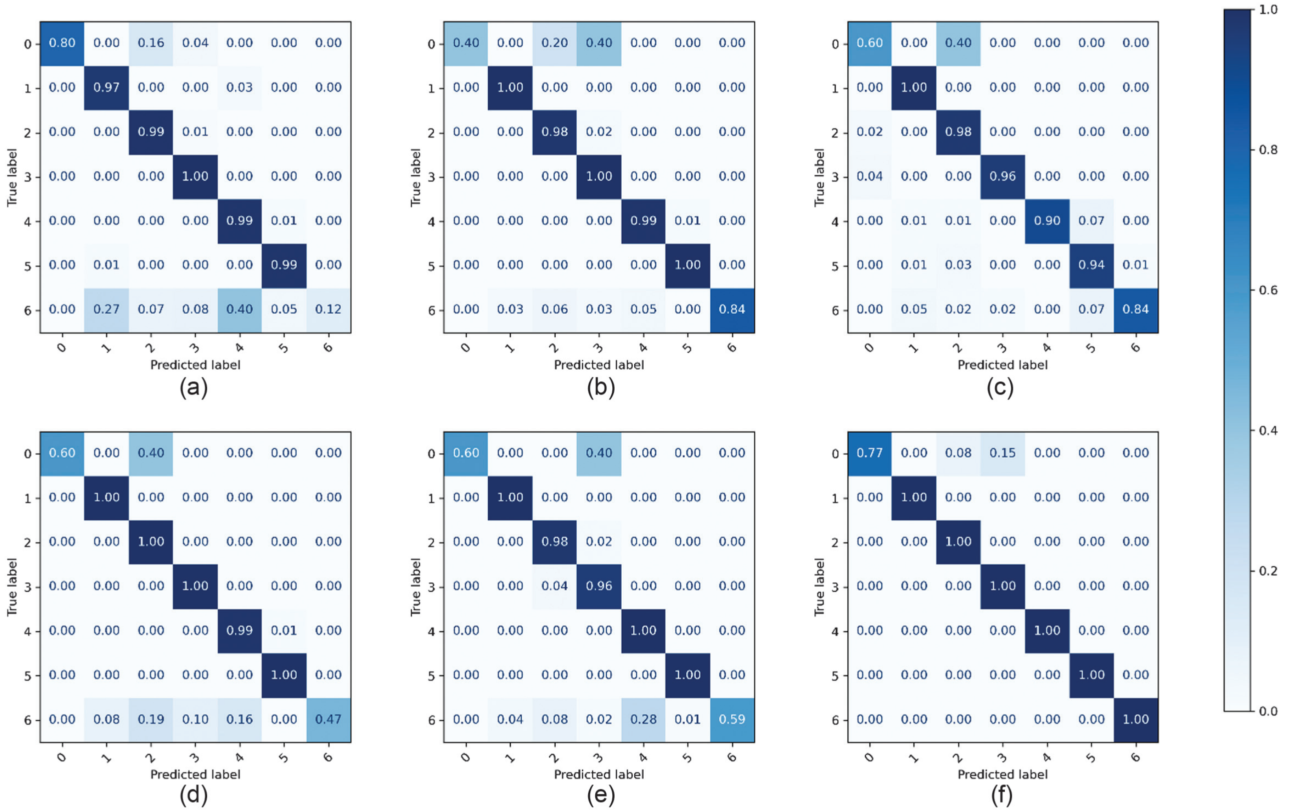

Fig. 7. Confusion matrices of fault classification results across comparative methods: (a) BiLSTM, (b) CNN, (c) GRU, (d) LSTM, (e) MLP, and (f) GCMLP.

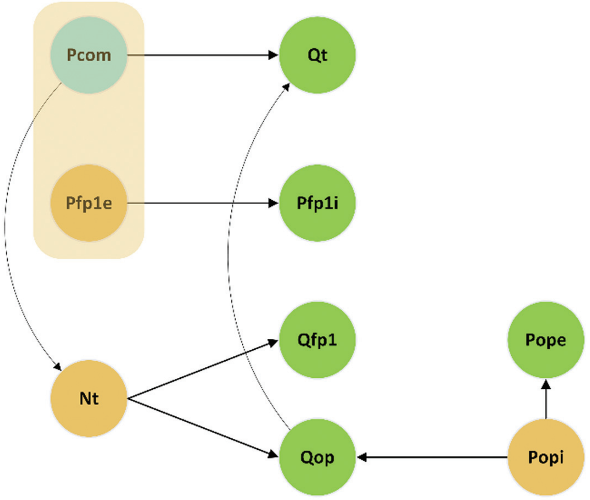

Fig. 8. GC diagram of variables in throat ablation process of LRE thrust chamber.

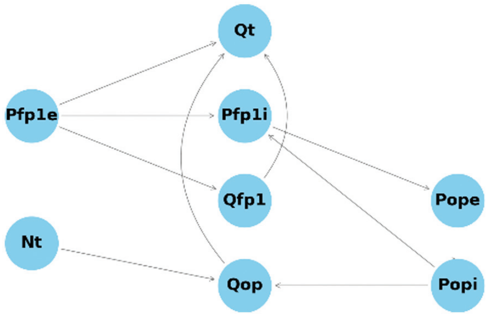

Fig. 9. GC-derived mapping of critical variables in LRE thrust chamber throat ablation from GCMLP.

speed N_{tur}. These causal relationships confirm P_{fp1e} and N_{tur} as exogenous variables.

(2) The coaxial integration of the main turbine, oxidizer pump, and fuel stage-1 pump forces synchronous rotational speed elevation. This speed synchronization increases flow rates in both the oxidizer pump and fuel stage-1 pump. Concurrently, reduced fuel stage-1 pump outlet pressure P_{fp1e} diminishes the booster turbine's driving capability, thereby lowering the fuel stage-1 pump inlet pressure. Furthermore, per mass conservation principles: oxidizer pump flow predominantly enters the main turbine through the gas generator, while fuel flow remains unchanged under regulator control. Consequently, oxidizer flow variations directly modify main turbine flow characteristics.

(3) The oxidizer booster pump derives power from the main turbine's exhaust gas. Within the gas generator, increased oxidizer flow combined with constant fuel flow elevates the mixture ratio. This mixture enrichment lowers gas temperature, subsequently reducing the temperature of the exhaust gas driving the oxidizer booster turbine. The diminished thermal energy impairs the oxidizer booster pump's head generation, decreasing the oxidizer pump's inlet pressure, consequently impacting oxidizer flow dynamics and outlet pressure. Since sensors cannot monitor all variables in this process, the propagation path is incompletely illustrated and represented through exogenous variables encapsulating latent factors.

Table IV. Fault classification accuracy of different sparsity penalty

Sparsity	Precision	Recall	Accuracy
Group lasso	89.90%	91.30%	91.78%
General group lasso	89.90%	91.30%	91.78%
Hierarchical group lasso	96.98%	97.51%	97.67%
Group sparse group lasso	93.03%	94.71%	95.50%

Based on the obtained GC graph, we approximate the direct and indirect influences between nearly equivalent variables during comparison and equivalize adjacent variables in the LRE system according to Equation 14 and 15. It is observed that the GCMLP can extract GC matrices characterized by separability and interpretability, which largely align with the actual GC values and fundamentally correspond to the underlying fault mechanisms. Therefore, GCMLP demonstrates its ability to enhance the credibility of diagnostic results by extracting physically meaningful features that endow the diagnostic model with interpretability.

C. DISCUSSION

As noted earlier, sparsity preservation is required throughout the GC extraction process. To explore how different sparsity constraints influence fault localization accuracy, an additional experiment was conducted. The experiment primarily evaluated the following sparsity penalties: group lasso penalty, general group lasso penalty, hierarchical group lasso penalty, and group sparse group lasso. Quantitative analysis includes metrics for fault classification accuracy, including precision, recall, and accuracy. Comparative results are summarized below:

It can be observed that the fault localization accuracy of GCMLP is 91.78% when using group lasso and general group lasso as sparsity losses to constrain GC extraction. When hierarchical group lasso and group sparse group lasso are employed as sparsity losses, the fault localization accuracies increase to 96.67% and 95.50%, respectively. Through analyzing different lasso penalty computation approaches and GC matrices, we can observe that group lasso groups together elements of the GC matrix with the same index across different lags, ensuring sparsity of the GC matrix in the lag dimension. General group lasso groups the GC effects of other variables on a specific variable, ensuring sparsity of the matrix in the variable dimension. In contrast, group sparse group lasso and hierarchical group lasso simultaneously address the sparsity of the GC matrix across both different lags and different variables, which allows them to concurrently obtain a sparse set of GC time series and a subset of relevant lags. Therefore, group sparse group lasso and hierarchical group lasso are more beneficial for extracting GC related to LRE and facilitating fault diagnosis. When designing new sparsity constraints, sparsity across multiple dimensions should be considered.

V. CONCLUSION

This study constructs an Encoder–Decoder architecture with dedicated feature extractors for distinct LRE variables. Specifically, we design a Component-wise MLP-based GC Extractor to derive GC relationships as diagnostic features. The correctness of GC extraction is validated through next-time-step variable prediction accuracy, while its discriminative power is verified via fault localization performance. Key results are summarized as follows:

(1) This study established a novel fault localization methodology that utilizes an Encoder–Decoder-structured MLP network to extract GC relationships among variables in LREs, achieving over 95% localization accuracy through GC-based diagnostic features.

(2) By analyzing GC interactions extracted by GCMLP in specific fault scenarios, we demonstrated strong conformance between the identified causal patterns and fundamental engine operational principles and fault propagation mechanisms, thereby providing physically interpretable diagnosis results.

(3) The proposed approach surpasses widely implemented models, including MLP, CNN, and time series optimized architectures (GRU, BiGRU, LSTM, and BiLSTM)—in fault localization accuracy while maintaining robust performance across varied sparsity penalty implementations.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

REFERENCES

[1] H. W. Lee *et al.*, “Satellite constellation pattern optimization for complex regional coverage,” *J. Spacecr. Rockets*, vol. 57, no. 6, pp. 1309–1327, 2020.

[2] M. Thangavelu, “USC ARTEMIS Project: Maximum Impact Moon Mission(MAXIM) Tribute to Apollo,” AIAA 2020–4098. ASCEND 2020. November 2020.

[3] S. Yang *et al.*, “Application issues of data-driven intelligent fault diagnosis technologies for liquid rocket engines,” *Acta Aeronautica Astronaut. Sin.*, vol. 46, no. 15, p. 131427, 2025.

[4] Y. Guo *et al.*, “Progress and development considerations on fault diagnosis techniques for solid rocket motor,” *J. Solid Rocket Technol.*, vol. 45, no. 1, pp. 4–12, 2022.

[5] S. Kanso *et al.*, “Remaining useful life prediction with uncertainty quantification of liquid propulsion rocket engine combustion chamber,” *IFAC-PapersOnLine*, vol. 55, no. 6, pp. 96–101, 2022.

[6] K. Kawatsu *et al.*, “Model-based fault diagnostics in an electromechanical actuator of reusable liquid rocket engine,” AIAA 2020–1624. AIAA Scitech 2020 Forum. January 2020.

[7] J. Cha *et al.*, “Fault detection and diagnosis algorithms for transient state of an open-cycle liquid rocket engine using nonlinear Kalman filter methods,” *Acta Astronaut*, vol. 163, pp. 147–156, 2019.

[8] L. Xu *et al.*, “Fault diagnosis of liquid rocket engine based on unscented kalman filter,” *Manned Spaceflight*, vol. 30, no. 4, pp. 516–525, 2024.

[9] R B. Sun *et al.*, “Fault sensor isolation method for liquid rocket engines based on multi-model,” *Mech. Syst. Sig. Process.*, vol. 225, p. 112278, 2025.

[10] B. Sun and C. Tian, “The fault real-time monitoring method for engine based on RMS value of characteristic frequency band,” *J. Rocket Propul.*, vol. 45, no. 4, pp. 74–78, 2019.

[11] C. Deng *et al.*, “Study on real-time diagnosis method of the main stage working condition of rocket engine based on improved ARMA Model,” *Comput. Meas &Control*, vol. 28, no. 2, pp. 33–38, 2020.

[12] X. Chen *et al.*, “Physics-informed deep neural network for bearing prognosis with multisensory signals,” *J. Dyn. Monit. Diagn.*, vol. 1, no. 4, pp. 200–207, 2022.

[13] Y. Wu *et al.*, “Fault diagnosis of linear guide rails based on SSTG combined with CA-DenseNet,” *J. Dyn. Monit. Diagn.*, vol. 3, no. 1, pp. 1–10, 2024.

[14] S. Y. Park and J. Ahn, “Deep neural network approach for fault detection and diagnosis during startup transient of liquid-propellant rocket engine,” *Acta Astronaut*, vol. 177, pp. 714–730, 2020.

[15] C. Wang *et al.*, “Dynamic model-assisted transferable network for liquid rocket engine fault diagnosis using limited fault samples,” *Reliab. Eng. Syst. Saf.*, vol. 243, p. 109837, 2024.

[16] Y. Yan *et al.*, Fault diagnosis method of launch vehicle based on sequential neural network[C]//2023 2nd International Symposium on Aerospace Engineering and Systems (ISAES). 2023: 208–212.

[17] C. W. J. Granger, “Investigating causal relations by econometric models and cross-spectral methods,” *Econometrica*, vol. 37, no. 3, pp. 424, 1969.

[18] A. Tank *et al.*, “Neural granger causality,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, 8, pp. 4267–4279, 2022.

[19] L. Ma, M. Wang, and K. Peng, “Nonlinear dynamic granger causality analysis framework for root-cause diagnosis of quality-related faults in manufacturing processes,” *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 3, pp. 3554–3563, 2024.

[20] X. Han *et al.*, Root Cause Analysis of Anomalies in Multivariate Time Series through Granger Causal Discovery[C]// The Thirteenth International Conference on Learning Representations. 2024.

[21] Z. Zhang and L. Wu, “Graph neural network-based bearing fault diagnosis using Granger causality test,” *Expert Syst. Appl.*, vol. 242, p. 122827, 2024.

[22] P. Tang, K. X. Peng, and J. Dong, “A novel method for deep causality graph modeling and fault diagnosis,” *Acta Autom. Sin.*, vol. 48, no. 6, pp. 1616–1624, 2022.

[23] C. Fan *et al.*, Interpretable Multi-Scale Neural Network for Granger Causality Discovery[C]//ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2023: 1–5.

[24] R. Marcinkevičs and J. E. Vogt, Interpretable Models for Granger Causality Using Self-explaining Neural Networks [A]. arXiv, 2021.

[25] P. Schwab, D. Miladinovic, and W. Karlen, “Granger-causal attentive mixtures of experts: Learning important features with neural networks,” *Proc AAAI Conf. Artif Intell*, vol. 33, no. 01, pp. 4846–4853, 2019.

[26] A. Fujita *et al.*, “Modeling gene expression regulatory networks with the sparse vector autoregressive model,” *BMC Syst Biol*, vol. 1, no. 1, p. 39, 2007.

[27] A. Shojaie and G. Michailidis, Discovering graphical Granger causality using the truncating lasso penalty.

[28] A. C. Lozano *et al.*, “Grouped graphical Granger modeling for gene expression regulatory networks discovery,” *Bioinformatics*, vol. 25, no. 12, pp. i110–i118, 2009.

[29] S. Basu and G. Michailidis, “Regularized estimation in sparse high-dimensional time series models,” *Annals Statistics*, vol. 43, no. 4, pp. 1535–1567, 2015.

[30] W. B. Nicholson, D. S. Matteson, and J. Bien, “VARX-L: Structured regularization for large vector autoregressions with exogenous variables,” *Int. J. Forecast.*, vol. 33, no. 3, pp. 627–651, 2017.

[31] J. V. Carreira, D. Costa, and J. G. Silva, “Fault injection spot-checks computer system dependability,” *IEEE Spectr.*, vol. 36, no. 8, pp. 50–55, 1999.