JRAIA INTERNATIONAL SYMPOSIUM 2025

Development of an AI/IoT-Based Air Conditioning Management System

○Hayato MORI*, Rinto KOZONO, Yusuke KAYASAKI**, Yuto KIMOTO* and Ryo TAKADA****

* FUJITSU GENERAL LABORATORIES LIMITED. 3-3-17, Suenaga, Takatsu-Ku, Kawasaki, 213-8502, Japan

** FUJITSU GENERAL LIMITED. 3-3-17, Suenaga, Takatsu-Ku, Kawasaki, 213-8502, Japan

ABSTRACT

This study aims to reduce HFC emissions and improve energy efficiency in commercial HVAC systems by enhancing maintenance diagnostics through a three-tiered approach. First, a nonlinear regression model based on refrigeration cycle simulations was developed to estimate minor refrigerant leakage from multiple operating parameters, enabling early detection of subtle anomalies. Second, a linear regression model using real operational logs was applied to learn correlations among key sensors such as temperature and pressure, detecting sensor faults through residual analysis. Third, a novel LLM-based framework was proposed, incorporating fault tree analysis (FTA) knowledge into prompts to perform zero-shot and few-shot reasoning for anomaly detection and root cause diagnosis in natural language. When sufficient training data are available, regression models achieve high accuracy, while the LLM approach complements them by reducing data preparation and retraining efforts. This hybrid strategy addresses key limitations of conventional diagnostics and demonstrates the feasibility of a rapid and flexible system, contributing to predictive maintenance and energy-efficient operation in commercial HVAC systems.

Keywords: Refrigerant leakage detection, Machine learning, Predictive maintenance

INTRODUCTION

As part of global efforts to mitigate climate change, reducing HFC emissions has become an international priority. Consequently, proper management of refrigerant leakage in commercial air-conditioning systems is a critical challenge. In addition, since HVAC systems account for a significant portion of a building's total energy consumption, preventing failures and maintaining or improving operational efficiency are essential for reducing environmental impact and lowering operating costs.

Traditionally, anomaly detection and fault diagnosis in HVAC systems have relied on machine learning using real operational data[1][2], as well as nonlinear regression analysis based on refrigeration cycle simulations or physical models[3][4][5]. While these methods can achieve high detection accuracy, they require significant effort for data collection and feature engineering, which limits their practicality when introducing new models or in environments with limited data availability. Furthermore, although the widespread adoption of IoT sensors has enabled real-time data acquisition, challenges remain in processing large volumes of data, adapting to diverse operating conditions, and reducing the burden of model retraining.

Recently, the zero-shot and few-shot reasoning capabilities of Large Language Models (LLMs), such as ChatGPT, have attracted attention. By leveraging domain knowledge as prompts, these models offer the potential to perform anomaly detection and root cause diagnosis without requiring large datasets. Based on this context, this study aims to enhance maintenance diagnostics for commercial HVAC systems through the parallel development and evaluation of three approaches:

1. Refrigerant Leakage Detection Using Nonlinear Regression
2. Sensor Anomaly Detection Using Linear Regression
3. LLM-Based Diagnostic Framework

DEVELOPMENT OF TECHNOLOGIES

APPROACH FOR VRF REFRIGERANT LEAKAGE DETECTION

The development of the VRF refrigerant leakage detection method began with the use of calculated data derived from a refrigeration cycle simulation model as the foundational dataset. Simulation enables the reproduction of a wide range of operating conditions, including rare scenarios and harsh environments that are difficult to capture through actual equipment testing. This approach allows the learning model to fit across diverse operational patterns, ensuring stable estimation within the interpolation range during real-world operation while mitigating output instability in extrapolated regions.

Simulations were executed to cover the operating condition ranges shown in Table 1, and the resulting refrigeration cycle data under various installation and operating environments were used as training data. Similar to prior studies (3–5), refrigeration cycle simulations were employed to generate the learning dataset.

Table 1. Operating Condition Ranges Used for Refrigeration Cycle Simulation

Parameter	Upper limit	Lower limit
Remaining Refrigerant (%)	100	40
Outdoor Air Temperature (°C)	46	20
Indoor Air Temperature (°C)	32	22
Connected Capacity Ratio (%) (Indoor Unit / Outdoor Unit)	150	50
Number of Operating Indoor Units	All units	1
Total Piping Length (m), Representative Example	A: 400	A: 35

SELECTION OF FEATURE VARIABLES IN MACHINE LEARNING

To develop a model applicable across multiple circuit types, five features were selected based on the relationship between refrigerant shortage-induced changes in the refrigeration cycle and sensor data commonly available in VRF systems:

- Subcooling degree at the outdoor heat exchanger
- Suction superheat degree
- Outdoor air temperature
- Pressure difference between discharge and suction sides
- Refrigerant mass flow rate at the compressor inlet

FEATURE ENGINEERING AND AI MODEL DEVELOPMENT

The proposed approach aims to leverage universal relationships between refrigerant shortage and refrigeration cycle behavior, using sensors typically installed in VRF systems. When refrigerant charge decreases, early-stage symptoms include a rise in suction pressure, a drop in discharge pressure, and reduced subcooling. As the shortage progresses, significant suction pressure reduction, increased suction superheat, and elevated discharge gas temperature are observed. These parameters can be monitored by standard VRF sensors and were therefore referenced when selecting input features for the AI learning algorithm.
For representative models A, B, and C, the following five common features were selected:

- Subcooling degree at the outdoor heat exchanger
- Suction superheat degree
- Outdoor air temperature
- Pressure difference between discharge and suction sides
- Refrigerant mass flow rate at the compressor inlet

The Extra Trees algorithm was adopted as the machine learning model, and hyperparameter optimization was performed using a grid search method.

VALIDATION
ACCURACY ON SIMULATION DATA

To verify the validity of the designed model, testing was conducted using refrigeration cycle simulation data. As shown in Fig. 1, the estimated results corresponded closely to the actual refrigerant shortage levels. This demonstrates that the proposed model can accommodate diverse piping lengths similar to those encountered in real installation environments.

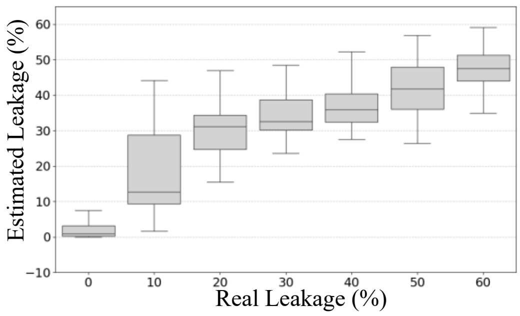

Fig. 1 Estimation results for the simulation of refrigerant leakage in VRF model A with long pipe length by the proposed model trained on standard pipe length and long pipe length

DETECTION USING ACTUAL OPERATIONAL DATA

Fig. 2 presents the estimation results of the developed model for a unit suspected of refrigerant leakage. The estimated refrigerant shortage was plotted on a daily basis. Signs of leakage were observed as early as 2019, with gradual progression over time. From 2021 onward, the model consistently estimated a shortage exceeding 30%. Notably, no customer complaints regarding cooling performance or operation were reported for this system; however, analysis of operational data by experienced HVAC engineers indicated signs of refrigerant shortage. This assessment aligned with the predictions generated by the developed model. These findings confirm that the proposed model can detect refrigerant leakage earlier than human observation.

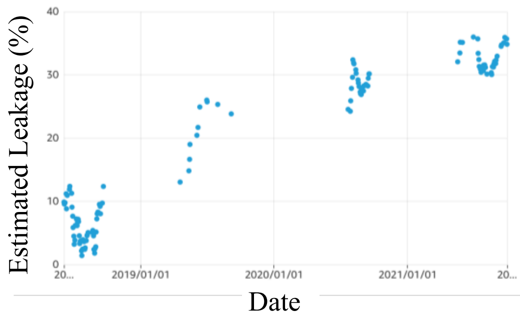

Fig. 2 Estimation results for the suspected leaks on actual equipment of VRF model B

SENSOR FAULT DETECTION

Efficient HVAC maintenance requires not only refrigerant leakage detection but also reliable identification of faults in key sensors and actuators. Ideally, such detection should leverage both operational and non-operational data. To this end, we developed a sensor fault detection technology using real-world VRF operational data, focusing on temperature and pressure sensors.

OVERVIEW OF THE SENSOR FAULT DETECTION ENGINE

The developed engine constructs a linear regression model to predict each target sensor value, assuming residuals follow a normal distribution. An observation is flagged as abnormal when the residual exceeds $\pm 3\sigma$ from the mean. The model uses outdoor air temperature as the explanatory variable and the target sensor value as the dependent variable, trained on operational data filtered to exclude periods when any compressor was running or immediately after shutdown, as these conditions introduce instability unrelated to outdoor temperature.

DATASET FOR EVALUATION

Evaluation was conducted using operational data from VRF outdoor units installed within the company. Details of the dataset are summarized in Table 2.

Table 2. Details of the Dataset Used for Evaluation

Item	Details
Number of Outdoor Units	149 units
Number of Sensors	Approximately 5–15
Period	2024/04/01 – 2025/03/31
Sampling Interval	1–5 minutes

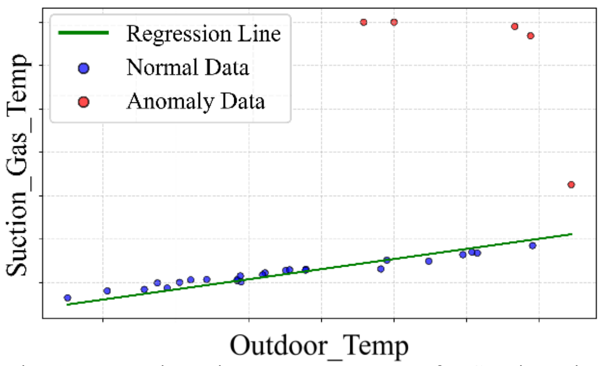

Fig. 3 Regression Line and Scatter Plot for Suction Pipe Temperature versus Outdoor Temperature

ANOMALY DETECTION AND EVALUATION

Fig. 3 illustrates the regression line and scatter plot for suction pipe temperature versus outdoor temperature. Normal data (blue) cluster near the regression line, while abnormal data (red) deviate significantly. Fig. 4 presents the residuals between predicted and observed values over time, confirming that anomalies were detected when residuals exceeded the threshold, while normal data remained stable.

Accuracy and false positive rate (FPR) were adopted as evaluation metrics. As shown in Table 3, all target sensors achieved an accuracy exceeding 95%, and the FPR remained as low as 3.9% at most. These results demonstrate that the proposed method provides sufficient accuracy for practical application.

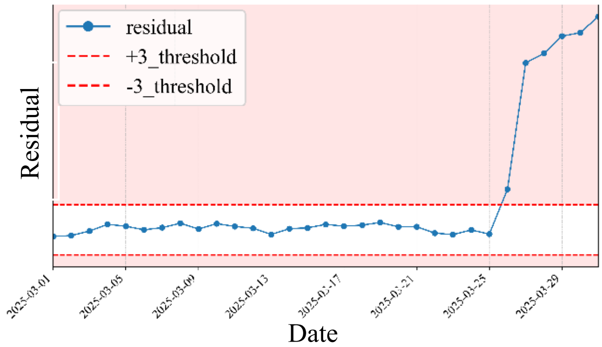

Fig. 4 Residual Analysis for Sensor Anomaly Detection

Table 3. Detection Accuracy and False Positive Rate for Each Sensor

Sensor	Accuracy	False Positive Rate (FPR)
Compressor 1	0.980	0.020
Compressor 2	0.994	0.006
Discharge Pipe 1	0.962	0.039
Discharge Pipe 2	0.992	0.008
Heat Exchanger Gas 1	0.968	0.032
Heat Exchanger Gas 2	0.962	0.038
Heat Exchanger Liquid 1	0.971	0.029
Heat Exchanger Liquid 2	0.970	0.030
Liquid Pipe 1	0.976	0.024
Liquid Pipe 2	0.979	0.021
SC Heat Exchanger Gas Inlet	0.967	0.033
SC Heat Exchanger Gas Outlet	0.989	0.011
Heat Exchanger Outlet	0.968	0.032

LLM – BASED ANOMALY DETECTION AND DIAGNOSIS

LLM APPROACH BACKGROUND

While regression models achieve high accuracy when sufficient labeled data are available, many field environments lack such data. Recent studies[6] [7] have shown that Large Language Models (LLMs) can perform zero-shot anomaly detection when domain knowledge is embedded in prompts. Building on these insights, we collaborated with service engineers to integrate Fault Tree Analysis (FTA) into an LLM-based diagnostic framework for VRF systems.

SYSTEM ARCHITECTURE

The proposed system (Fig. 5) comprises four layers: data preprocessing, knowledge integration, LLM inference, and post-processing. Inputs to the LLM include:

- Domain Knowledge: FTA-based structured tables linking anomalies, causes, evidence, and verification steps.
- Operational Data: Time-series logs and metadata (units, encoding).
- Diagnostic Prompt: Detailed instructions covering data interpretation, reasoning steps, and output format.

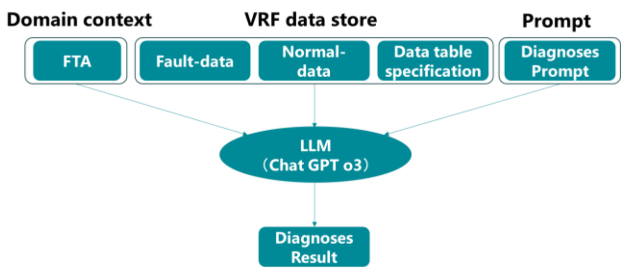

Fig. 5 Input/output configuration of anomaly diagnosis method using LLM

LANGUAGE MODEL SELECTION

Conventional generative LLMs are designed to produce statistically likely sequences of words based on input text, which does not necessarily make them suitable for deep causal reasoning. In contrast, reasoning-oriented LLMs incorporate internal chains of thought prior to response generation, enabling consistent decision-making for problems involving multiple interdependent factors. Given the complexity and diversity of fault causes in VRF systems, we concluded that adopting a reasoning-oriented model was essential.

To meet the requirements for zero-shot/few-shot learning, long-context processing, and high-speed API access, we selected OpenAI's o3 model, released on April 16, 2025.

DOMAIN KNOWLEDGE INTEGRATION

To enable accurate diagnostics, this study incorporated structured prior knowledge of potential VRF anomalies and their causes into the input data. Possible approaches for knowledge integration include:
(i) model fine-tuning,
(ii) retrieval-augmented generation (RAG), and
(iii) direct ingestion of structured knowledge files.

Considering development efficiency and ease of updates, we adopted approach (iii). Fault Tree Analysis (FTA) and knowledge graphs were normalized into tabular format and provided to the LLM as part of the input. Additionally, prompts instructed the model to "refer to the table and evaluate consistency during diagnosis," enabling structured reasoning. As a result, the system produced highly consistent anomaly diagnoses aligned with FTA-based causal analysis.

PROMPT DESIGN

To ensure that anomaly diagnosis using LLMs is executed as intended, it is essential to design prompts aligned with the diagnostic objectives. Even when the necessary data are provided, ambiguous prompts can lead to variability in the analysis process and reduced diagnostic accuracy.

Our approach explicitly specifies analysis steps, conditions, and precautions within the prompt to minimize variability. The analysis procedure consists of three main stages:

1. Loading and interpreting operational data,
2. Incorporating prior knowledge, and
3. Performing fault analysis.

For data loading, the prompt includes not only code execution instructions but also semantic interpretation steps to prevent data misuse or misinterpretation of prior knowledge. During fault analysis, the prompt provides detailed instructions on data cleansing, sensor and control values to verify, statistical metrics to compute, and comparison methods, ensuring that prior knowledge is applied appropriately. Incorporating domain expertise from technical teams into these instructions is critical.

We also adopted widely recognized prompt engineering techniques to enhance stability and accuracy. By combining these methods, the designed prompts enable consistent and reliable anomaly diagnosis using LLMs.

VALIDATION AND RESULTS

Experiments were conducted to evaluate the proposed method using actual VRF systems that experienced anomalies. Two types of anomalies were tested: (i) refrigerant shortage due to leakage and (ii) sensor failure. For refrigerant leakage, historical operational data were prepared for both cooling and heating modes when the system was operating under a 50% refrigerant shortage. The sensor failure case involved an abnormal suction temperature during the cooling season, where the thermistor consistently reported values exceeding 50 °C.

REFRIGERANT SHORTAGE DIAGNOSIS (COOLING MODE)

The diagnostic results output by the LLM are shown in Table 4. As indicated, the model correctly identified refrigerant shortage as the most probable anomaly. Supporting evidence included suction pressure drop and expansion valve opening, which are characteristic indicators of refrigerant shortage in operational data. The prompt design allows multiple diagnostic candidates to be output, requiring comparison of likelihood based on evidence. Among the candidates, refrigerant shortage was determined to be the most probable fault.

Table 4. Diagnostic results for refrigerant shortage cases (cooling mode).

Suspected Abnormality	Main Causes
Refrigerant Deficiency (Leakage) Probability: High	1. Significant drop in suction pressure 2. Electronic expansion valve remains excessively open 3. Compression ratio increases, but insufficient cooling capacity is observed 4. Even under identical indoor and outdoor load conditions, the operating parameters differ significantly
Electronic Expansion Valve Malfunction Probability: Medium	1. Suction pressure remains abnormally low despite no refrigerant shortage 2. The valve may not be actuating properly 3. Low suction pressure is not reflected in discharge pressure, which remains within the normal range
Suction Pressure Sensor Drift Probability: Low	1. Discrepancy between sensor readings and actual operating conditions 2. Poor electrical contact at the sensor terminal

REFRIGERANT SHORTAGE DIAGNOSIS (HEATING MODE)

Table 5 presents the diagnostic results for heating mode under refrigerant shortage conditions. The results confirm that the proposed method can also diagnose refrigerant shortage during heating operation. The model cited suction pressure drop, expansion valve opening, and discharge temperature rise as evidence, all of which are phenomena associated with refrigerant shortage. Similar to cooling mode, the diagnosis was accurate and consistent.

Table 5. Diagnostic results for refrigerant shortage cases (heating mode).

Suspected Abnormality	Main Causes
Refrigerant Deficiency (Leakage) Probability: High	1. Low suction pressure 2. Average electronic expansion valve opening is six times the normal level 3. Compression ratio and average discharge superheat are abnormally high 4. Combination of the above is a typical indication of refrigerant shortage
Electronic Expansion Valve Control Failure Probability: Medium	1. EEV remains fully closed 2. Even though suction pressure should decrease, it remains abnormally high, indicating valve malfunction 3. Suction pressure increases abnormally while discharge pressure does not reflect the same trend
Indoor Fan Malfunction Probability: Low	1. Airflow is significantly lower than expected 2. Possible explanation when temperature imbalance occurs across multiple indoor units

Table 6. Diagnostic results for sensor failure.

Suspected Abnormality	Main Causes
Refrigerant Deficiency / Insufficient Charge Probability: High	1. Significant increase in suction superheat, indicating a high superheat condition 2. Low-pressure side is not particularly low, but the electronic expansion valve remains wide open (indicating insufficient refrigerant flow) 3. Compressor current increases (due to high load), yet low suction pressure is observed, suggesting a “starved” operating condition
Electronic Expansion Valve Malfunction Probability: Medium	1. Valve may not be actuating properly, as suction pressure does not decrease despite control commands 2. Possible issue with internal step motor of the valve
Compressor Mechanical Failure Probability: Low	1. Even with high valve opening, suction pressure does not decrease 2. Discharge superheat is lower than normal by 2–3%, suggesting reduced compression efficiency, possibly due to internal leakage

SENSOR FAILURE DIAGNOSIS

Table 6 shows the diagnostic results for a sensor failure case. Ideally, the model should have identified a suction temperature thermistor fault; however, it incorrectly diagnosed refrigerant leakage and expansion valve malfunction. Although the abnormal rise in suction temperature was detected, concurrent changes in electronic expansion valve position and compressor frequency likely led to misclassification as refrigerant leakage. Other suggested anomalies included expansion valve and compressor malfunctions, but the model did not explicitly identify sensor failure as the root cause.

DISCUSSION

The proposed method successfully extracted key indicators—such as expansion valve opening and low-pressure symptoms—from operational data under refrigerant shortage conditions, enabling accurate diagnosis of refrigerant deficiency. These diagnostic results were consistent with FTA-based prior knowledge, confirming that the analysis was appropriately performed.

However, in cases involving sensor anomalies, the method tended to misinterpret secondary effects as primary faults. In VRF systems, thermistor failures can alter control values and propagate inconsistencies throughout the system, making it difficult for the LLM to isolate the true root cause. This indicates that LLM reasoning alone is insufficient for robust diagnostics in all scenarios.

To address this challenge, it is considered effective to supplement LLM inference with complementary approaches such as statistical anomaly detection and rule-based methods that explicitly account for control specifications. Such a hybrid strategy is expected to reduce misclassification, enhance diagnostic robustness and explainability, and align more closely with practical field requirements where reliability, transparency, and ease of deployment are critical.

CONCLUSION

This study advanced maintenance diagnostics for VRF air-conditioning systems through a three-layered approach: nonlinear regression with simulation data, linear regression with operational logs, and FTA-enhanced LLM diagnostics.

The findings indicate that, in building a practical maintenance diagnostic support system, it is not sufficient to rely on a single method. Instead, it is considered effective to apply different techniques appropriately and combine them where necessary—for example, regression models for data-rich environments, LLM + FTA for data-scarce cases, and hybrid approaches that integrate LLM reasoning with rule-based or statistical checks for sensor-related anomalies.

Overall, the results suggest that such a hybrid diagnostic strategy can improve accuracy, reduce misclassification, and better meet the requirements of real-world HVAC maintenance, where reliability and practicality are critical.

References

[1] H. Kato, M. Ito, H. Hokari, K. Higashi and M. Ishigaki, “Current Status and Future Outlook of Refrigerant Leakage Detection Technologies”, JRAIA International Symposium, (2023).

[2] M. Yoshimi, S. Kasahara, T. Hikawa, S. Yamada, N. Noda, K. Kuroda and S. Kamiaka, “Development of a Remote Refrigerant Leakage Detection System Using Machine Learning Technique”, JRAIA International Symposium, (2023)

[3] S. Okumura, H. Ueda and N. Shiraishi, “AI Service Development by Machine Learning using IoT Data and Simulation Data”, Proceedings of the Annual Conference of JSAI, (2020). (in Japanese)

[4] S. Sasaki, S. Okumura and H. Ueda, “Determining the application domain of machine learning using anomaly detection and application to refrigerant leak detection”, Proceedings of the Annual Conference of JSAI, (2022). (in Japanese)

[5] S. Sasaki and H. Ueda, “DEVELOPMENT OF REFRIGERANT LEAK DETECTION TECHNOLOGY FOR AIR CONDITIONERS USING IOT TECHNOLOGY” , JRAIA International Symposium, (2023).

[6] S. Alnegheimish, L. Nguyen, L. Berti-Equille and K. Veeramachaneni, “Large language models can be zero-shot anomaly detectors for time series?”, arXiv preprint arXiv:2405.14755, (2024).

[7] A. Russell-Gilbert, A. Sommers, A. Thompson, L. Cummins, S. Mittal, S. Rahimi, M. Seale, J. Jaboure, T. Arnold and J. Church, “AAD-LLM: Adaptive Anomaly Detection Using Large Language Models”, arXiv preprint, arXiv:2411.00914, (2024).