TU Delft

Delft University of Technology

Document Version
Final published version

Licence
Dutch Copyright Act (Article 25fa)

Citation (APA)
Lu, C., & Itard, L. (2026). Leveraging LLM for P&ID-based Automated Code Generation in HVAC Fault Detection and Diagnosis. In C. Zilio, F. Busato, L. Mazzarella, & M. Noro (Eds.), *Proceedings of the 15th REHVA HVAC World Congress - CLIMA 2025* (pp. 399-407). (Lecture Notes in Civil Engineering; Vol. 788 LNCE). Springer. https://doi.org/10.1007/978-3-032-10546-2_37

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.



Leveraging LLM for P&ID-based Automated Code Generation in HVAC Fault Detection and Diagnosis

Chujie Lu(✉) and Laure Itard

Department of Architectural Engineering and Technology, Delft University of Technology, Delft, Netherlands
cj.lu@tudelft.nl

Abstract. Fault detection and diagnosis (FDD) play a crucial role in minimizing energy waste and reducing maintenance costs in HVAC systems. Diagnostic Bayesian networks (DBNs), as probabilistic graphical models, offer a promising solution due to robustness to uncertainties, modeling flexibility, scalability, and interpretability. However, the current DBN construction is either a tedious and time-consuming manual process or relies heavily on training data, posing significant barriers to wide-spread adoption. This study proposes a novel large language model (LLM)-driven framework for automating DBN code generation for HVAC systems by extracting knowledge from process and instrumentation diagrams (P&IDs), extending beyond the reliance on traditional sensor data. The results demonstrate that the proposed framework can generate functional DBN code, reasonable symptoms, and DBN parameters. However, fault diagnosis experiments revealed that only the "supply fan stuck" fault was correctly identified, underscoring the need for further refinement. Future work will focus on enhancing LLM capabilities, such as prompt engineering and fine-tuning, and optimizing DBN parameters using limited data to improve diagnostic accuracy.

Keywords: Large Language Model · Code Generation · Fault Detection and Diagnosis · Bayesian Network · HVAC System

1 Introduction

Buildings account for approximately 40% of energy consumption in the European Union and over one-third of energy-related greenhouse gas emissions. The revised Energy Performance of Buildings Directive aims to cut emissions in the building sector by at least 60% by 2030 (compared to 2015 levels) and achieve climate neutrality by 2050 [1]. Achieving these goals requires minimizing energy waste, particularly in heating, ventilation, and air conditioning (HVAC) systems, which are the largest energy consumers in buildings. However, HVAC systems are prone to faults in sensing, mechanical, and control components, leading to indoor discomfort, poor air quality, and excessive energy consumption. Therefore, developing automatic fault detection and diagnosis (FDD) methods for HVAC systems is essential for effectively reducing energy waste and maintenance costs [2].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
C. Zilio et al. (Eds.): CLIMA 2025, LNCE 788, pp. 399–407, 2026.
https://doi.org/10.1007/978-3-032-10546-2_37

FDD methods for HVAC systems have been extensively studied, which can broadly be categorized into two types: knowledge-based and data-driven [3, 4]. Knowledge-based FDD methods rely on predefined rules and are the most widely used in the existing HVAC systems, but they often suffer from low accuracy issues. Over the past decade, data-driven FDD methods have become popular, including support vector machine, artificial neural network, random forest, and Transformer. They leverage statistical analysis and machine learning to automatically identify faulty patterns, often achieving high diagnostic accuracy. However, data-driven FDD methods require high-quality labeled faulty datasets for model training, which can be very time-consuming and costly to collect [5]. Also, their reliance on training data poses challenges related to generalization and scalability [6]. Additionally, many data-driven FDD methods are considered as black-box models, can lead to skepticism regarding their reliability and understanding [7].

To address these challenges, diagnostic Bayesian networks (DBNs) present a promising solution due to their alignment with HVAC design and operation practices [8]. They provide several advantages for FDD in HVAC systems, including robustness to uncertainties, modeling flexibility, scalability, and interpretability. DBNs have been successfully applied to FDD for individual HVAC components, such as chillers, boilers, and pumps, as well as aggregated systems like air handling units (AHUs) [9–11], variable air volume (VAV) terminals [12], and heat pumps [13]. Moreover, they have been utilized in diagnosing faults across entire HVAC networks and even district energy systems [14, 15]. However, constructing DBNs for HVAC systems involves several challenges in practice. One of the main difficulties is that the current DBN construction is either a tedious and time-consuming manual process [9, 10, 16], especially as HVAC systems become increasingly complex and large, or relies heavily on training data, similar to data-driven methods, posing significant barriers to widespread adoption.

Recent advancements in large language models (LLMs) present new opportunities to overcome these barriers. On the one hand, LLMs, pre-trained by large-scale real-world data, have demonstrated a strong ability to understand and apply domain knowledge, including concepts relevant to HVAC systems and fault diagnosis [17]. On the other hand, LLMs possess powerful code generation and auto-correction capabilities across various engineering domains, such as chemical engineering, software engineering, and building energy engineering [18]. By leveraging these strengths, LLMs have the potential to streamline the traditionally labor-intensive DBN development process into automated construction, making FDD solutions more accessible for real-world HVAC applications.

This study proposes a novel LLM-driven framework for automating DBN code generation for HVAC systems by extracting knowledge from process and instrumentation diagrams (P&IDs), extending beyond the reliance on traditional sensor data. The proposed framework is evaluated to diagnose faults in an AHU system of a Dutch office building. The rest of the paper is organized as follows. Sect. 2 briefly introduces DBNs, the proposed LLM-driven framework of automating DBN code generation for HVAC systems, and the case systems. Sect. 3 provides the information on the case study. Sect. 3 presents the preliminary results of the proposed framework. Sect. 4 discusses future work. Sect. 5 concludes this paper.

2 Methodology

2.1 Diagnostic Bayesian Network

DBNs are probabilistic graphical models that utilize directed acyclic graphs to represent random variables and their conditional dependencies. In the context of FDD for HVAC systems, In the context of FDD in building energy systems, represent a set of possible faults and represent a set of possible symptom states. If a fault F_i happens when a symptom S is observed or assumed to have occurred, the probability of the fault F_i under the symptom S can be written as follows.

$$P(F_i|S) = \frac{P(F_iS)}{P(S)} = \frac{P(S|F_i)P(F_i)}{P(S)} \tag{1}$$

where is the posterior probability of the fault F_i given symptom S; is the joint probability; is the conditional probability of observing symptom S given that fault F_i has occurred; is the prior probability of the fault F_i; $P(S)$ is the marginal probability of observing the symptom S, which can be expressed as follows.

$$P(S) = \sum_{i=1}^{n} P(S|F_i)P(F_i) \tag{2}$$

To systematically and effectively construct DBN for HVAC systems, the four symptoms and three faults (4S3F) reference architecture is proposed, illustrated in Fig. 1, which incorporates three generic types of faults, i.e. model, component, and control faults, and four generic types of symptoms, i.e. balance, energy performance, operational state, and additional symptoms.

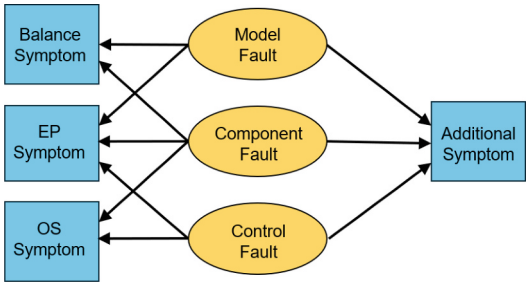


Fig. 1. The illustration of 4S3F reference architecture

2.2 LLM-driven Framework of Automating DBN Code Generation

Fig. 2 presents the proposed LLM-driven framework of automating DBN code generation for HVAC systems. The aim is to leverage LLM to extract the knowledge from P&ID and automatically construct DBNs tailored to specific HVAC systems. The proposed framework consists of the following key steps.

First, transform HVAC P&ID diagrams into machine-readable formats such as JSON, Industry Foundation Classes (IFC), or xgXML [19, 20]. Since LLMs primarily process textual input, these structured formats allow the LLM to systematically extract component location, sensor dependencies, and system information, forming a foundation for accurate DBN construction.

Second, input the structured P&ID text into the LLM using carefully designed prompts to guide the model in generating DBN code. Several popular LLMs can be utilized for this task, which can be categorized into open-source and closed-source models. Popular open-source LLMs include BERT, LLaMA, T5, and DeepSeek. Closed-source models include GPT-3.5, GPT-4, and Claude [17]. Moreover, prompt engineering is crucial for guiding the LLMs in accurately interpreting HVAC system components and generating relevant DBN structures [21].

Third, assess the generated code quality to ensure its correctness, efficiency, and applicability. This includes performing syntax checks to verify the code is free from errors. Additionally, fault symptom verification is conducted to assess if the generated DBN accurately represents HVAC system behavior, especially in identifying fault pathways.

Last, validate the generated DBNs using real-world or simulated data to assess their FDD performance.

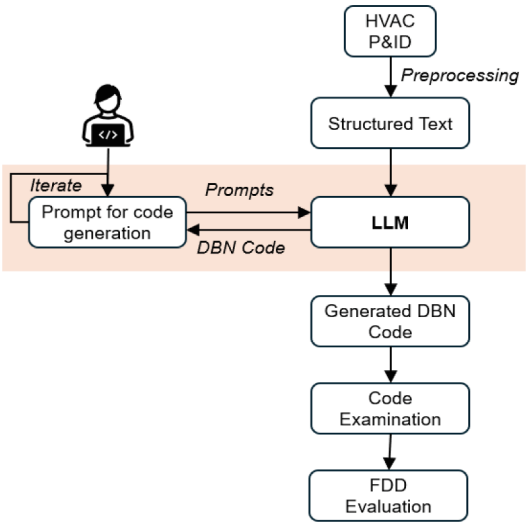


Fig. 2. LLM-driven framework of automating DBN code generation for HVAC systems

2.3 Case Study

This study utilizes a building manage system (BMS) dataset from the Kropman office building in Breda, the Netherlands, which is a living lab of the Brains4Building project. The case building has an approximate floor area of 1500 m^2. This study mainly focuses

on the central AHU system under heating mode, equipped with one heat recovery wheel (HRW) and heating coil. The simplified P&ID of the AHU system is illustrated in Fig. 3. The BMS variables are listed in Table 1. More information about the case study can be found in [11, 16].

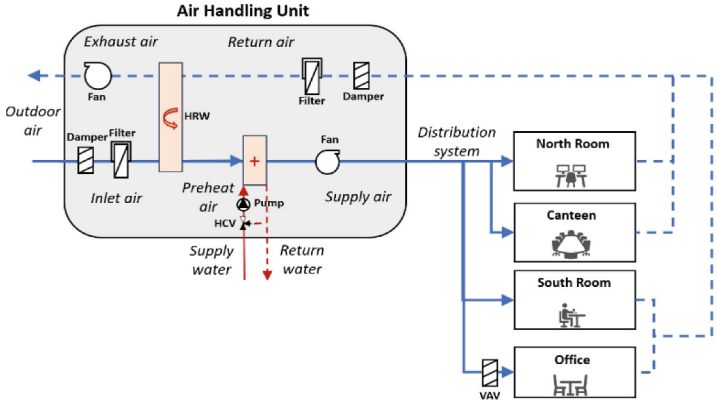


Fig. 3. The simplified P&ID of the AHU in the case building

Table 1. Variables in the BMS

Variable	Description	Variable	Description
Ts	Supply air temperature	*Ps*	Supply air pressure
To	Outdoor air temperature	*PdT1*	Supply air filter differential pressure
RHo	Outdoor air related humidity	*PdT2*	Return air filter differential pressure
Tp	Preheated air temperature	*PdT3*	Supply fan differential pressure
RHp	Preheated air related humidity	*Tsw*	Supply water temperature
Tr	Return air temperature	*Trw*	Return water temperature
RHr	Return air related humidity	*Wfan*	Supply fan power
Te	Exhaust air temperature	*Tset*	Supply air temperature setpoint
RHe	Exhaust air related humidity	*Pset*	Supply air pressure setpoint

In this study, JSON is chosen as the structured text format for its simplicity and ease of integration. Claude 3.5 Sonnet, developed by Anthropic, is selected as the LLM considering its strong capabilities in natural language understanding, reasoning, and code generation. Chain of Thought (CoT) [21] is selected as the prompt method to enhance the performance of LLMs in generating accurate DBN code for HVAC systems. Referring to the generic modeling procedure of DBNs in building energy systems, CoT can encourage the LLM to explicitly reason through the problem step-by-step, simulating human-like

thought processes. The generated DBN is required to diagnose three common faults in AHUs: “Heat Recovery Wheel (HRW) Stuck”, “Heating Coil Valve (HCV) Stuck”, and “Supply Fan Stuck”. The prompting is presented in Fig. 4.

Prompt
Generate Python code using pgmpy to build a Diagnostic Bayesian Network (DBN) for fault diagnosis in an Air Handling Unit (AHU) within an HVAC system.

Let's break the task down step-by-step:
1. Understand the AHU Components ("PID.txt") and BMS Variables ("BMS.txt"). Control strategies: The supply air pressure setpoint is fixed at 250 Pa. A Proportional-Integral (PI) controller regulates the heating coil valve position based on supply air temperature and its setpoint. The Heat Recovery Wheel (HRW) has a higher priority for regulating temperature than the heating coil.
2. Define fault nodes: "Heat Recovery Wheel Stuck", "Heating Coil Valve Stuck", and "Supply Fan Stuck".
3. Define the related symptoms (rules) using the BMS variables.
4. Define the prior probabilities of faults and conditional probabilities
5. Inference and fault diagnosis.

Fig. 4. The CoT Prompting in this study

3 Preliminary Result

3.1 Code Examination

First, the correctness of the generated code is verified. In the initial trial of code generation using the CoT prompting method, a compilation issue was encountered. There were inconsistencies between the DBN structure and the conditional probabilities distribution (CPD) definitions. Specifically, the error was “ValueError: CPD associated with ‘Supply_Temp_Error’ doesn’t have proper parents associated with it.” To resolve this, the error was fed back into the LLM, which then adjusted the structure. The LLM made three key changes to ensure that the DBN structure now properly reflects the causal relationships:

- Removed the direct connection from “HC_Valve_Fault” to “Supply_Temp_Error” (it now goes through “HC_DT”).
- Removed the direct connection from “Fan_Fault” to “Supply_Pressure_Error” (it now goes through “Fan_DP”).
- Added the missing connection from “HRW_Efficiency” to “Supply_Temp_Error”.

In short, LLM generated functional DBN code for FDD.

Then, the faults and symptoms, DBN structure, and DBN parameters are checked. In the generated code, four key symptoms were identified to assess potential HVAC faults:

- “Supply_Temp_Error”: it calculates the absolute difference between *Ts* and *Tset*. If this deviation exceeds 2 °C, the symptom is detected.
- “Supply_Pressure_Error”: it calculates the absolute difference between Ps and the expected setpoint (250 Pa). If the deviation exceeds 25 Pa, the symptom is detected.
- “HRW_Efficiency”: the actual HRW efficiency is calculated as (*Te-To*)/(*Tr-To*). If this efficiency drops below 0.6, the symptom is detected.

- "Fan_DP": if the PdT3 is below 100 Pa, the symptom is detected.

Based on expert knowledge, the generated symptoms are reasonable and mentioned in related literature [9–11].

The prior probabilities of three faults being "present" were set as 5%, indicating a low initial likelihood of failure, which is reasonable. The conditional probabilities varied based on symptom dependencies.

3.2 Fault Diagnosis Results

To evaluate the FDD performance of the automatically generated DBN, three fault experiments were performed in the case building. The experiment description and diagnostic results are presented in Table 2. The fault diagnosis rule is that a fault is diagnosed if it has the highest posterior probability and exceeds a diagnostic threshold of 15%.

Table 2. The fault posterior probabilities in the fault experiments

No.	Description	Fault Nodes			Diagnostic Result
		HRW Stuck	HCV Stuck	Supply Fan Stuck	
1	Supply fan speed stuck at 30%	10.34%	1.35%	**48.6%**	Correct
2	HCV openness stuck at 10%	51.24%	8.2%	0.55%	Fail
3	HRW rotation speed stuck at 30%	0.43%	3.81%	0.55%	Fail

In Case 1, the fault "Supply Fan Stuck" was correctly diagnosed with the highest posterior probability with 48.6%. the supply fan speed was stuck at 30%, which was too low, triggering the symptom "Fan_DP" due to insufficient differential pressure. In Case 2, the fault "HRW Stuck" was wrongly diagnosed with 51.24%. This misdiagnosis occurred because the symptom "HRW_Efficiency" was incorrectly detected, likely due to inaccurately estimated efficiency thresholds. Notably, these thresholds were determined by the LLM rather than derived from in-situ measurements, which likely have led to discrepancies in symptom detection. In Case 3, all the posterior probabilities were low and there was no detected evidence.

4 Discussion and Future Work

The results show that the proposed LLM-driven framework can generate functional DBN code. Future research will focus on several key areas to enhance the accuracy and applicability of the proposed LLM-driven DBN generation framework.

Evaluating various LLMs (e.g., GPT-3.5, GPT-4, Llama, and more) alongside advanced prompt engineering strategies such as few-shot learning and retrieval-augmented generation can further optimize DBN generation accuracy. Additionally, fine-tuning on domain-specific HVAC data may enhance model performance.

Developing adaptive DBN by leveraging small-scale real-world data for refining thresholds and CPDs. P&IDs can provide only the basic DBN structure and fault-symptom relationships. To improve diagnostic accuracy, it is crucial to incorporate limited real-world commissioning data for refining thresholds and CPDs. Therefore, it is necessary to collaborate with certain amounts of commissioning data to adjust DBNs.

Extending the framework to diagnose a wider range of HVAC faults and applying it to larger, more complex systems (e.g., multi-zone air handling units and district heating/cooling networks) will increase its practical applicability in real-world scenarios.

5 Conclusions

This study proposes a novel LLM-driven framework for automating DBN code generation for HVAC systems by extracting knowledge from P&IDs, extending beyond the reliance on traditional sensor data. By leveraging these strengths, LLMs have the potential to streamline the traditionally labor-intensive DBN development process into automated construction, making FDD solutions more accessible for real-world HVAC applications. There are four key steps in the proposed framework: transforming HVAC P&ID diagrams into machine-readable formats, inputting the structured P&ID text into the LLM using carefully designed prompts, assessing the generated code quality, and validating the generated DBNs. The results demonstrate that the proposed framework can generate functional DBN code, reasonable symptoms, and DBN parameters. However, fault diagnosis results show that only the "supply fan stuck" fault was correctly identified, highlighting the need for further refinement, including deeper investigation into LLM capabilities and parameter optimization to improve diagnosis accuracy.

Acknowledgement. This study is a part of the Brains4Buildings project, sponsored by the Dutch grant program for Mission-Driven Research, Development, and Innovation (MOOI). The authors would like to thank Srinivasan Gopalan (TU Eindhoven), Karzan Mohammed (TU Eindhoven), and Dr. Shalika Walker (Kropman) for data collection and collaboration.

References

1. Maduta, C., D'Agostino, D.: Readiness of Zero-Emission Buildings (ZEBs) implementation in the European Union. In: Noro, M., Zilio, C. (eds.) E3S Web of Conferences, p. 04009 (2024)
2. Gopalan, S., Rijs, A., Chitkara, S., Thamban, A., Kramer, R.: Fault prioritisation for Air Handling units using fault modelling and actual fault occurrence data. Energ. Buildings. **319**, 114476 (2024)
3. Chen, J., Zhang, L., Li, Y., Shi, Y., Gao, X., Hu, Y.: A review of computing-based automated fault detection and diagnosis of heating, ventilation and air conditioning systems. Renew. Sust. Energ. Rev. **161**, 112395 (2022)
4. Zhao, Y., Li, T., Zhang, X., Zhang, C.: Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future. Renew. Sust. Energ. Rev. **109**, 85–101 (2019)
5. Taheri, S., Ahmadi, A., Mohammadi-Ivatloo, B., Asadi, S.: Fault detection diagnostic for HVAC systems via deep learning algorithms. Energ. Buildings. **250**, 111275 (2021)

6. Lu, C., Li, S., Lu, Z.: Building energy prediction using artificial neural networks: A literature survey. Energ. Buildings. **262**, 111718 (2022)
7. Heimar Andersen, K., Pommerencke Melgaard, S., Johra, H., Marszal-Pomianowska, A., Lund Jensen, R., Kvols Heiselberg, P.: Barriers and drivers for implementation of automatic fault detection and diagnosis in buildings and HVAC systems: An outlook from industry experts. Energ. Buildings. **303**, 113801 (2024)
8. Lu, C., Wang, Z., Mosteiro-Romero, M., Itard, L.: Diagnostic Bayesian network in building energy systems: Current insights, practical challenges, and future trend. Energ. Buildings. **341**, 115845 (2025)
9. Zhao, Y., Wen, J., Xiao, F., Yang, X., Wang, S.: Diagnostic Bayesian networks for diagnosing air handling units faults – part I: Faults in dampers, fans, filters and sensors. Appl. Therm. Eng. **111**, 1272–1286 (2017)
10. Zhao, Y., Wen, J., Wang, S.: Diagnostic Bayesian networks for diagnosing air handling units faults - Part II: Faults in coils and sensors. Appl. Therm. Eng. **90**, 145–157 (2015)
11. Wang, Z., Lu, C., Taal, A., et al.: Bayesian network-based fault detection and diagnosis of heating components in heat recovery ventilation. In: E3S Web Conf., vol. 672, p. 07022 (2025)
12. Xiao, F., Zhao, Y., Wen, J., Wang, S.: Bayesian network based FDD strategy for variable air volume terminals. Autom. Constr. **41**, 106–118 (2014)
13. Cai, B., et al.: Multi-source information fusion based fault diagnosis of ground-source heat pump using Bayesian network. Appl. Energy. **114**, 1–9 (2014)
14. Taal, A., Itard, L., Zeiler, W.: A reference architecture for the integration of automated energy performance fault diagnosis into HVAC systems. Energ. Buildings. **179**, 144–155 (2018)
15. Taal, A., Itard, L.: P&ID-based automated fault identification for energy performance diagnosis in HVAC systems: 4S3F method, development of DBN models and application to an ATES system. Energ. Buildings. **224**, 110289 (2020)
16. Lu, C., Wang, Z., Mosteiro-Romero, M., Itard, L.: Introducing causality to symptom baseline estimation: A critical case study in fault detection of building energy systems. In: Proceedings of a Sim Conference 2024: 5th Asia Conference of IBPSA, pp. 1621–1628. Osaka, Japan (2024)
17. Lu, J., Tian, X., Zhang, C., Zhao, Y., et al.: Evaluation of large language models (LLMs) on the mastery of knowledge and skills in the heating, ventilation and air conditioning (HVAC) industry. Energy Built Environ. **6**(5), 875–892 (2024)
18. Jiang, G., Ma, Z., Zhang, L., Chen, J.: EPlus-LLM: A large language model-based computing platform for automated building energy modeling. Appl. Energy. **367**, 123431 (2024)
19. Chamari, L., Petrova, E., Pauwels, P.: An End-to-End Implementation of a Service-Oriented Architecture for Data-Driven Smart Buildings. IEEE Access, vol. 11, pp. 117261–117281 (2023)
20. Andriamamonjy, A., Saelens, D., Klein, R.: An automated IFC-based workflow for building energy performance simulation with Modelica. Autom. Constr. **91**, 166–181 (2018)
21. Wei, J., et al.: Chain-of-thought prompting elicits reasoning in large language models. In: Advances in Neural Information Processing Systems, vol. 35, pp. 24824–24837 (2022)