scientific reports

OPEN

Transforming free-text coronary angiography reports into structured, analyzable data using large language models

Ji Woo Song[1], Ji Yong Jang[2], Hyeongsoo Kim[2], Young-Guk Ko[3] & Seng Chan You[4,5]✉

Coronary angiography (CAG) reports contain many details about coronary anatomy, lesion characteristics, and interventional procedures. However, their free-text format limits their research utility. Therefore, we sought to develop and validate a framework leveraging large language models (LLMs) to convert CAG reports automatically into a standardized structured format. Using 50 CAG reports from a tertiary hospital, we developed a multi-step framework to standardize and extract key information from CAG reports. First, a standard annotation schema was developed by cardiologists. Thereafter, an LLM (GPT-4o) converted the free-text CAG reports into the hierarchical annotation schema in a standardized format. Finally, clinically relevant information was extracted from the standardized schema. One hundred CAG reports from each of two hospitals were used for internal and external test, respectively. The 12 key information points included four CAG-related (previous stent information, lesion characteristics, and anatomical diagnosis) and eight percutaneous coronary intervention (PCI)-related key points (complex PCI criteria and current stent information). For internal test, two interventional cardiologists independently extracted information, with discrepancies resolved through consensus, as reference standard. Based on the reference standard, the proposed framework demonstrated superior accuracy for CAG-related (99.5% vs. 91.8%; $p < 0.001$) and comparable accuracy for PCI-related key points (98.3% vs. 97.4%; $p = 0.512$) in the internal test. External test confirmed high accuracy for both CAG- (96.2%) and PCI-related key points (99.4%). This framework demonstrated excellent accuracy in standardizing free-text CAG reports, potentially enabling more efficient utilization of detailed clinical data for cardiovascular research.

Keywords Artificial intelligence, Coronary angiography, Database, Percutaneous coronary intervention, Standardization

Coronary angiography (CAG) reports contain information about coronary artery disease and percutaneous coronary intervention (PCI), providing invaluable data for clinical research. However, CAG reports are limited by their unstructured, free-text format, which makes it difficult to search, analyze, and process data consistently[1]. With continued growth in the secondary use of clinical data, observational studies on coronary artery diseases using electronic health records (EHR) continue to increase; however, many of these studies are conducted without the use of detailed information on complex coronary anatomy or procedures[2–4]. Moreover, the lack of structured data also hampers the identification of eligible patients for clinical trials based on specific anatomical or procedural criteria[5], and the quality assessment or performance monitoring of interventional procedures[6].

To standardize complex medical information in free-text medical records, Park et al. have previously proposed Staged Optimization of Curation, Regularization, and Annotation of clinical text (SOCRATex). This framework initially requires domain experts to define a standardized schema that specifies how clinical information should be organized. Thereafter, experts manually review clinical notes and annotate relevant information according

[1]Yonsei University College of Medicine, Seoul, Korea. [2]Division of Cardiology, National Health Insurance Service Ilsan Hospital, Goyang, Korea. [3]Department of Cardiology, Severance Hospital, Yonsei University College of Medicine, Seoul, Korea. [4]Yonsei Institute for Digital Health, Yonsei University, Seoul, South Korea. [5]Dssepartment of Biomedical Systems Informatics, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea. ✉email: chandryou@yuhs.ac

to this predefined schema to create standardized, machine-readable data[7]. Although this systematic approach effectively converts unstructured clinical notes into analyzable data, it involves a time-consuming manual annotation process and requires significant expert involvement, hampering its large-scale implementation[5].

Recent advancements in large language models (LLMs), such as ChatGPT, a type of artificial intelligence (AI) used for understanding and processing free-text, can structure vast amounts of free-text within EHRs with minimal programming effort[8]. Recent reviews have demonstrated that LLMs consistently outperform rule-based and earlier machine learning–based approaches, such as BERT, for structuring unstructured medical narratives owing to their superior generalization to diverse phrasing[9]. In fact, various studies have reported using LLMs to convert free-text radiology and pathology reports into structured formats, highly accurately[10–12]. Within cardiovascular imaging, however, LLM applications have predominantly focused on either non-invasive coronary CT angiography (e.g., CAD-RADS standardization)[13,14] or right heart catheterization[15,16], while the application of LLMs to transform CAG reports—with their complex procedural details—into machine-readable formats remains unexplored.

Therefore, building upon the previous work on hierarchical annotation, this study sought to develop a two-step framework that converts free-text CAG reports into a hierarchical, machine-readable schema using an LLM, then applies rule-based logic to extract 12 predefined key points. We then validated the framework internally against cardiologist extraction and externally on an independent hospital set, and we test generalizability using alternative LLMs with identical prompts.

Results

Dataset characteristics

Table 1 summarizes the composition of the training, internal test, and external test datasets according to the year of procedure and the type of included coronary intervention. The training dataset consisted of 50 reports, including 26 combined CAG + PCI, 21 CAG-only, 1 PCI-only, and 2 other reports (angiography of coronary artery bypass grafts). The internal test dataset included 100 reports composed of 41 CAG + PCI, 55 CAG-only, 3 PCI-only, and 1 other report. The external test dataset comprised 100 reports, with 57 CAG + PCI, 41 CAG-only, 1 PCI-only, and 1 other report (angiography of inferior vena cava filter placement). Across all datasets, both CAG- and PCI-related information were often intermingled within the same report.

Internal test

Table 2 and Fig. 1 show the accuracy of the framework using GPT-4o and the mean accuracy of the two cardiologists in extracting clinical key points from CAG reports in the internal test dataset.

The framework demonstrated a significantly higher mean accuracy than that of the cardiologists in extracting the four CAG-related key points (99.5% vs. 91.8%, $p < 0.001$). For individual CAG key points, the framework showed superior accuracy in *location of previous stents* (100.0% vs. 91.1%, $p = 0.001$), *previous stent information* (100.0% vs. 95.3%, $p = 0.032$), and *location and type of lesion* (97.9% vs. 80.7%, $p < 0.001$). Both the framework and cardiologists achieved 100% accuracy for *anatomical diagnosis*.

For the eight PCI-related key points, the framework showed comparable accuracy to the cardiologists across all items, with mean accuracy scores of 98.3% versus 97.4% ($p = 0.512$). The framework and the cardiologists performed similarly for each specific point: *multivessel PCI* (100.0% vs. 100.0%), *≥ 3 lesions treated* (97.7% vs. 97.7%), *bifurcation PCI with ≥ 2 stents* (100.0% vs. 97.7%), *≥ 3 stents implanted* (97.7% vs. 97.7%), *CTO PCI* (100.0% vs. 100.0%), *total stent length > 60 mm* (97.7% vs. 97.7%), *complex PCI* (97.7% vs. 98.9%), and *current stent information* (95.5% vs. 89.8%). All p values were above 0.05, indicating no significant differences between the framework and the cardiologists. More detailed results are shown in Supplementary Table 1.

A qualitative analysis was done to understand the significant discrepancies in accuracy between the framework using GPT-4o and the cardiologists. As shown in Supplementary Table 2, cardiologists often generated errors

Characteristics of the training, internal test and external test set			
	Training set	Internal test set	External test set
Total number of reports	50	100	100
Year of procedure			
2009–2013	10	25	0
2014–2018	22	44	0
2019–2023	18	31	0
2024	0	0	100
Included procedure			
CAG and PCI	26	41	57
CAG only	21	55	41
PCI only	1	3	1
Others	2*	1*	1†

Table 1. Characteristics of the training, internal test, and external test datasets. *CAG* coronary angiography, *PCI* percutaneous coronary intervention.*Reports about angiography of coronary artery bypass graft; †Report about inferior vena cava filter placement.

Key information points	Framework	Cardiologists*	*p* value
Total 4 CAG-related key information points	99.5% (98.1 to 99.9)	91.8% (89.6 to 93.6)	< 0.001
Location of previous stents	100.0% (96.2 to 100.0)	91.1% (86.2 to 94.8)	0.001
Previous stents information	100.0% (96.2 to 100.0)	95.3% (91.3 to 97.8)	0.032
Location and type of lesion	97.9% (92.7 to 99.7)	80.7% (74.4 to 86.1)	< 0.001
Anatomical diagnosis	100.0% (96.2 to 100.0)	100.0% (98.1 to 100.0)	1
Total 8 PCI-related key information points	98.3% (96.3 to 99.4)	97.4% (96.0 to 98.5)	0.512
Multivessel PCI	100.0% (92.0 to 100.0)	100.0% (95.9 to 100.0)	1
≥ 3 lesions treated	97.7% (88.0 to 99.9)	97.7% (92.0 to 99.7)	1
Bifurcation with ≥ 2 stents	100.0% (92.0 to 100.0)	97.7% (92.0 to 99.7)	0.552
≥ 3 stents implanted	97.7% (88.0 to 99.9)	97.7% (92.0 to 99.7)	1
CTO PCI	100.0% (92.0 to 100.0)	100.0% (95.9 to 100.0)	1
Total Stent Length > 60 mm	97.7% (88.0 to 99.9)	97.7% (92.0 to 99.7)	1
Complex PCI	97.7% (88.0 to 99.9)	98.9% (93.8 to 100.0)	1
Current stents information	95.5% (84.5 to 99.4)	89.8% (81.5 to 95.2)	0.335

Table 2. Accuracy percentages with 95% confidence intervals for the analysis of 12 key information points by the framework and by two cardiologists in the internal test. *The independent extractions performed by the two cardiologists have been combined into a single group and the mean accuracy was calculated. *CTO* chronic total occlusion.

by mislabeling the location of previous stents, current stents, and lesions. They also frequently misclassified the type of lesions due to missed key descriptors, such as "eccentric" and "ostium". In contrast, the framework consistently performed precise extraction of information from very complex cases, as shown in Supplementary Fig. 3A and 3B.

Multi-model results

When the same framework was applied using Gemini-2.5-Flash and Claude-4.5-Sonnet, both models achieved comparable or superior performance to the cardiologists in analyzing 12 key information points of CAG and PCI. Specifically, for the 4 CAG-related key information points, both models showed significantly higher accuracy than cardiologists ($p < 0.001$), whereas for the 8 PCI-related key information points, the performances were non-inferior ($p > 0.05$). The detailed accuracy percentages and confidence intervals are summarized in Supplementary Table 3.

External test

Table 3 shows the accuracy of the framework in extracting clinical key points from CAG reports in the external test dataset. For the four CAG-related key points, the framework achieved a mean accuracy of 96.2%, showing consistently high performance across each item. For the eight PCI-related key points, the framework achieved a mean accuracy of 99.4%, also showing consistently high performance across each item. These results indicated the framework's robust accuracy in extracting key details from both CAG and PCI data. Supplementary Table 4 shows more detailed results.

A qualitative analysis was conducted to understand the reasons for errors made by the framework. As shown in Supplementary Table 5, some inevitable errors arose from the inherent ambiguity and inconsistencies within the CAG section of the reports themselves. However, the PCI sections of the reports were structured using seven key items for each treated lesion—"Location," "Guiding catheter," "Guidewire," "Preballoon," "Adjuvant balloon," "DEB," and "Stent"—which provided a well-organized foundation for data extraction. This pre-existing structure in the free-text report significantly facilitated the conversion into a machine-readable format by the LLM, resulting in high accuracy for extracting PCI-related key points. Supplementary Fig. 4 depicts two representative cases from the external test set, showcasing the excellent capability of the framework in converting all the details of a highly complex CAG and PCI case into a machine-readable format.

Discussion

This study shows that a two-step approach—LLM standardization followed by rule-based extraction—can transform free-text CAG reports into analyzable data with accuracy comparable to or exceeding cardiologists on CAG key points and non-inferior performance on PCI key points, with consistent results in an external cohort and across alternative LLMs. These results suggest that automated standardization of complex medical documents is not only feasible, but can be implemented with high reliability, potentially transforming how we utilize clinical information embedded in unstructured medical reports.

To our knowledge, no previous study has attempted to convert detailed information about coronary anatomy and catheterization procedures from free-text reports into a machine-readable format. A recent systematic review indicated that natural language processing research in the field of cardiology remains underexplored as compared to that in the field of oncology[17]. Although the availability of coded information in EHRs and claims data has increased substantially, most large-scale observational studies have not utilized detailed information about coronary anatomy or procedural characteristics[18]. While studies leveraging data from dedicated registries

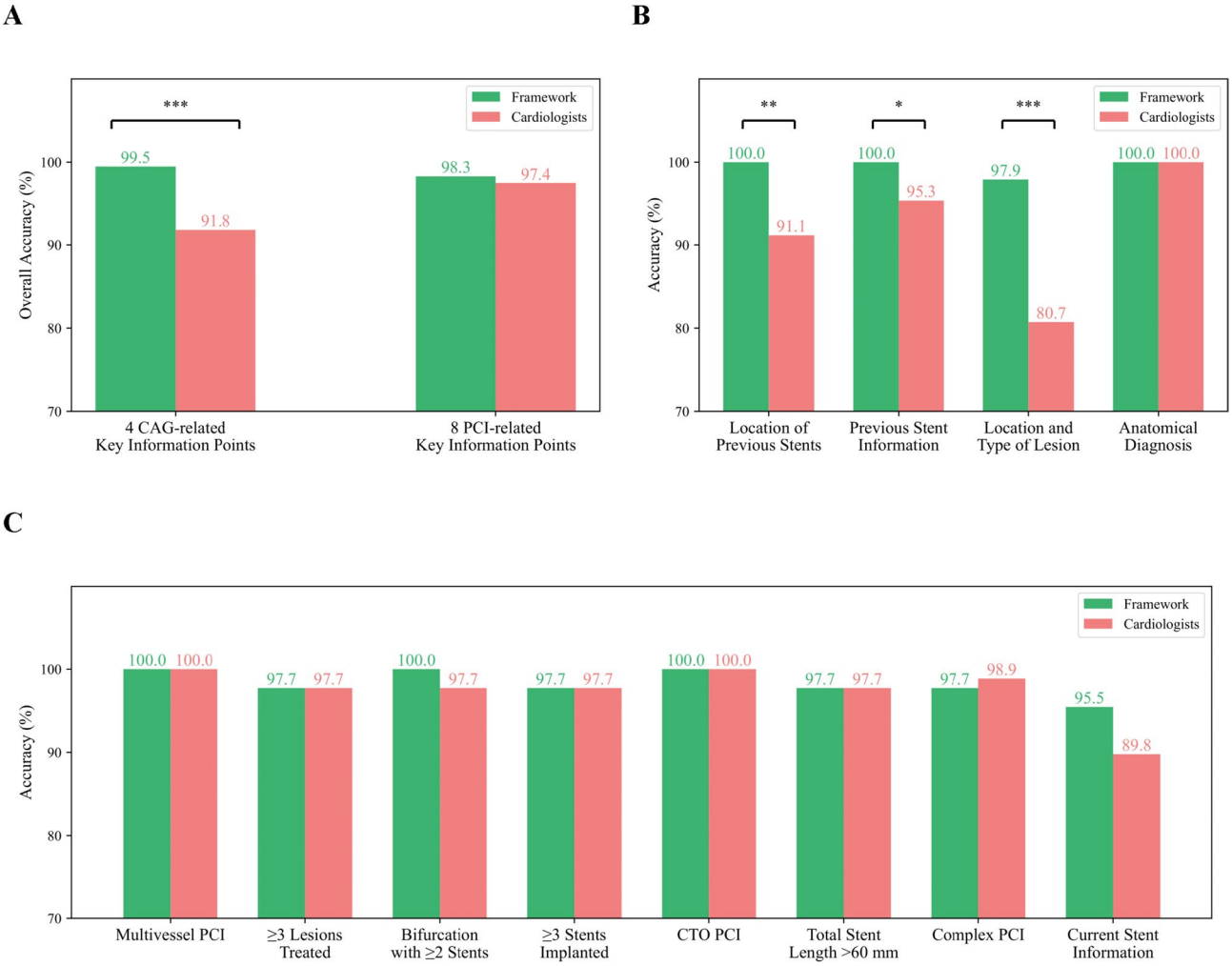

Fig. 1. Bar plot showing the accuracy percentages with 95% confidence intervals for the analysis of 12 key information points by the framework and by two cardiologists in the internal validation process. This figure compares the accuracy of the framework and of the two cardiologists in analyzing 12 key information points during the internal validation process. (**A**) shows the overall accuracy for four coronary angiography (CAG)-related and eight percutaneous coronary intervention (PCI)-related key points. (**B**) presents individual accuracy for the four CAG-related key points. (**C**) displays the accuracy for the eight individual PCI-related key points. Statistical significance is indicated as follows: *for $p < 0.05$, **for $p < 0.01$, and ***for $p < 0.001$.

have incorporated such information, this approach requires marked manual effort. Our framework addresses this limitation by automatically converting detailed procedural and anatomical information into a structured, machine-readable format, which promotes several critical capabilities: (1) systematic analysis of large-scale coronary intervention outcomes across different patient populations and institutions, (2) efficient quality assessment and performance monitoring of interventional procedures, (3) facilitation of clinical research by enabling rapid identification of eligible patients for trials based on specific anatomical or procedural criteria[19], and (4) providing support for clinical decision-making by making historical procedural data more accessible and analyzable.

The framework developed in this study uses a two-step approach to ensure accuracy and interpretability. In a previous study, while a one-step approach using LLMs was able to capture basic information accurately, it demonstrated difficulty with complex medical reasoning[11]. For example, in pathology reports, LLMs could accurately identify tumor measurements but had difficulty determining accurate cancer staging. To overcome this limitation, we separated our process into two distinct steps: first, using LLMs to organize the basic information from CAG reports into an expert-defined standardized format, and second, applying specific rules developed by cardiologists to build clinically relevant data. This approach, similar to methods successfully used in pathology research[12], combines the LLM's strength in understanding medical text with cardiologist-designed rules for interpreting nuance and complexity in coronary anatomy and catheterization. Furthermore, an instruction prompt was given to LLM to encode domain-specific knowledge and detailed rules, as guided by cardiologists, to ensure accurate and consistent mapping of the intricate information contained in CAG reports.

Key information points	Accuracy of framework
Total 4 CAG-related key information points	96.2%
Location of previous stents	95.9%
Previous stents information	98.0%
Location and type of lesion	94.9%
Anatomical diagnosis	95.9%
Total 8 PCI-related key information points	99.4%
Multivessel PCI	100.0%
≥ 3 lesions treated	100.0%
Bifurcation with ≥ 2 stents	98.3%
≥ 3 stents implanted	100.0%
CTO PCI	100.0%
Total stent length > 60 mm	100.0%
Complex PCI	100.0%
Current stents information	96.6%

Table 3. Accuracy percentages for the analysis of 12 key information points by the framework in the external validation test. *CTO* chronic total occlusion.

The significance of our work extends beyond its immediate performance metrics. By developing and publicly releasing a comprehensive hierarchical annotation schema for CAG reports, we have provided a standardized framework that other institutions can readily adopt or modify for their specific needs. In addition, the framework proved to be model-agnostic when tested with alternative large language models. Using the same internal test dataset and identical prompts, we re-evaluated the framework with Gemini-2.5-Flash and Claude-4.5-Sonnet. Both models achieved results comparable to those obtained with GPT-4o, showing significantly higher accuracy than cardiologists for CAG-related key information points ($p < 0.001$ for both) and non-inferior performance for PCI-related key points ($p > 0.05$). These findings (Supplementary Table 2) underscore that the proposed hierarchical annotation framework can maintain high reliability across different LLMs, demonstrating its generalizability beyond a single model implementation. This flexibility and generalizability are particularly important as healthcare institutions increasingly seek to process sensitive clinical data by using internal systems rather than external services. With the rapid advance in the capabilities of open-source LLMs, our framework offers a practical blueprint for standardization of automated medical documents that could be implemented using various LLM options, while maintaining high accuracy and reliability. The primary advantage is that it saves time and human effort.

The study had some limitations. First, only CAG reports from hospitals in South Korea were used to validate the robustness of the framework. To ensure thorough validation, the framework should be implemented on CAG reports from more diverse countries. Second, this study utilized only proprietary LLMs as the LLM for standardization. Therefore, further testing is needed to determine whether its high accuracy and reliability will be maintained even with less-competent open-source models, such as Llama-3[20]. Third, unlike the internal validation, the external validation process only evaluated the framework's accuracy as judged by cardiologists, without direct comparison to cardiologists' manual reviews. Although the framework achieved excellent accuracy in the external test, a direct comparison with manual reviews would have more clearly demonstrated its capabilities. Fourth, in both the internal and external validation processes, accuracy was evaluated based on only a single output from the LLM. Although we set the LLM's "temperature"—a parameter that controls the randomness and variability of the model's responses—to zero, in order to maximize predictability, slight variability remained due to the model's inherent characteristics. Fifth, this study did not directly evaluate the downstream clinical utility of the proposed framework. Although it demonstrated high accuracy in transforming CAG reports into structured data, we did not conduct follow-up analyses to assess its impact on real-world research tasks—such as enabling large-scale observational studies or facilitating the identification of eligible patients for clinical trials. Future work should therefore focus on applying this framework to actual cardiovascular datasets to evaluate its practical contribution to clinical research and decision support.

In conclusion, we developed a novel framework for standardizing free-text CAG reports and extracting complex clinical data. This framework demonstrated excellent accuracy in standardizing CAG reports, indicating its potential for more efficient utilization of detailed clinical data in these reports for cardiovascular research. Future research should focus on applying this framework to real-world cardiovascular datasets to assess its effectiveness in accelerating large-scale observational studies and improving the efficiency of clinical trial eligibility identification based on free-text CAG reports.

Methods

Study design and datasets

CAG studies of adults (> 18-years-old), performed from January 1, 2009, to December 31, 2023 (15 years), at Severance Hospital, a tertiary hospital, were retrieved. From these, 50 and 100 CAG reports were randomly selected for the training and internal test sets, respectively. Following the Society for Cardiovascular Angiography and Interventions' expert consensus,[6] these CAG reports included the coronary anatomy and described lesions

with their location, severity, and morphological characteristics. PCI-related information, including the types and sizes of catheters, balloons, stents, and adjuvant devices, was thoroughly detailed in the CAG report. All reports were written in English, with rare exceptions of Korean texts describing emergency situations during procedures.

Furthermore, 100 random CAG reports from the National Health Insurance Service Ilsan Hospital, a secondary hospital, obtained between January 1, 2023, and December 31, 2023, were retrieved as an external test set.

This study was approved by the Institutional Review Board (IRB) of Severance Hospital (IRB No. 2024-1928-002), Seoul, Korea, and the IRB of the National Health Insurance Service Ilsan Hospital (IRB No. 2024-09-006). The requirement for informed patient consent was waived by the IRBs due to the retrospective nature of the study.

Framework overview

Figure 2 illustrates our CAG report standardization framework. After development of a hierarchical annotation schema by experience cardiologists (detailed below), the framework consisted of standardization and extraction steps. In the standardization step, an LLM converted free-text CAG reports into a standardized format based on the expert-defined schema. Thereafter, an automated extraction algorithm extracted the required clinical information. We intentionally adopted the two-step design—LLM-based standardization followed by rule-based extraction—to leverage the LLM's strength in text normalization while ensuring interpretability and consistency through deterministic, cardiologist-defined rules. This hybrid design mitigates errors observed in prior one-step LLM extraction approaches for complex medical reasoning tasks, providing both scalability and clinical transparency[11]. Figure 3 details the development and validation of the framework, including iterative refinement with interventional cardiologists to ensure clinical accuracy and reliability.

Development of the hierarchical annotation schema

A domain-specific hierarchical annotation schema was collaboratively developed by interventional cardiologists (JYJ) and clinical informatics experts (JWS and SCY). This schema defined how to organize key clinical information from CAG reports systematically, including coronary anatomy, lesion characteristics, and procedural details. Through iterative testing using the training dataset, we refined the schema to ensure comprehensive and accurate capture of complex clinical information. The schema was implemented in JavaScript Object Notation (JSON)

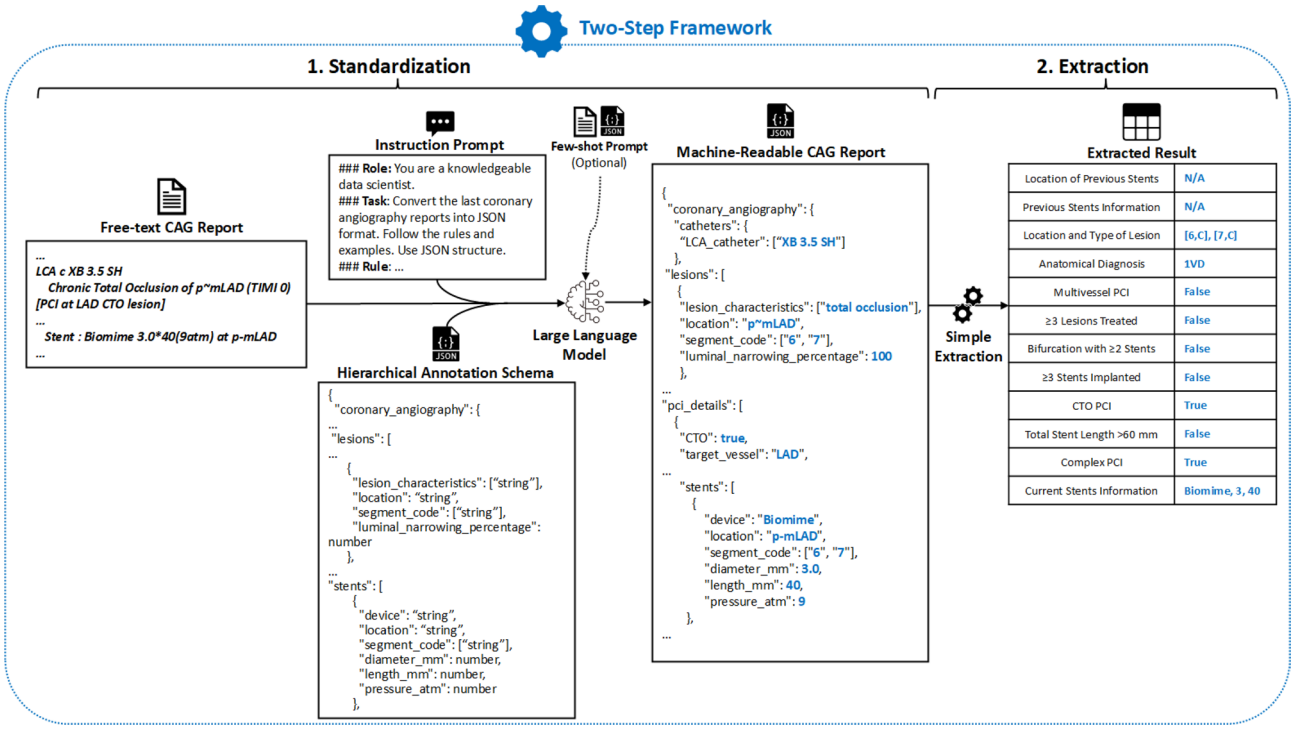

Fig. 2. Overview of the two-step framework for free-text CAG report standardization and data extraction. The framework consists of two primary steps: (1) standardization and (2) extraction. In the standardization step, free-text coronary angiography (CAG) reports are converted into a machine-readable hierarchical annotation schema using large language models (LLMs), guided by instruction prompts and optional few-shot prompts. This schema defines the structure and content of the machine-readable format generated in this manner. In the extraction step, 12 key data points are extracted from these machine-readable reports to enable automated and precise data analysis, providing insights into clinical variables, such as lesion characteristics, stent details, and complex percutaneous coronary intervention information.

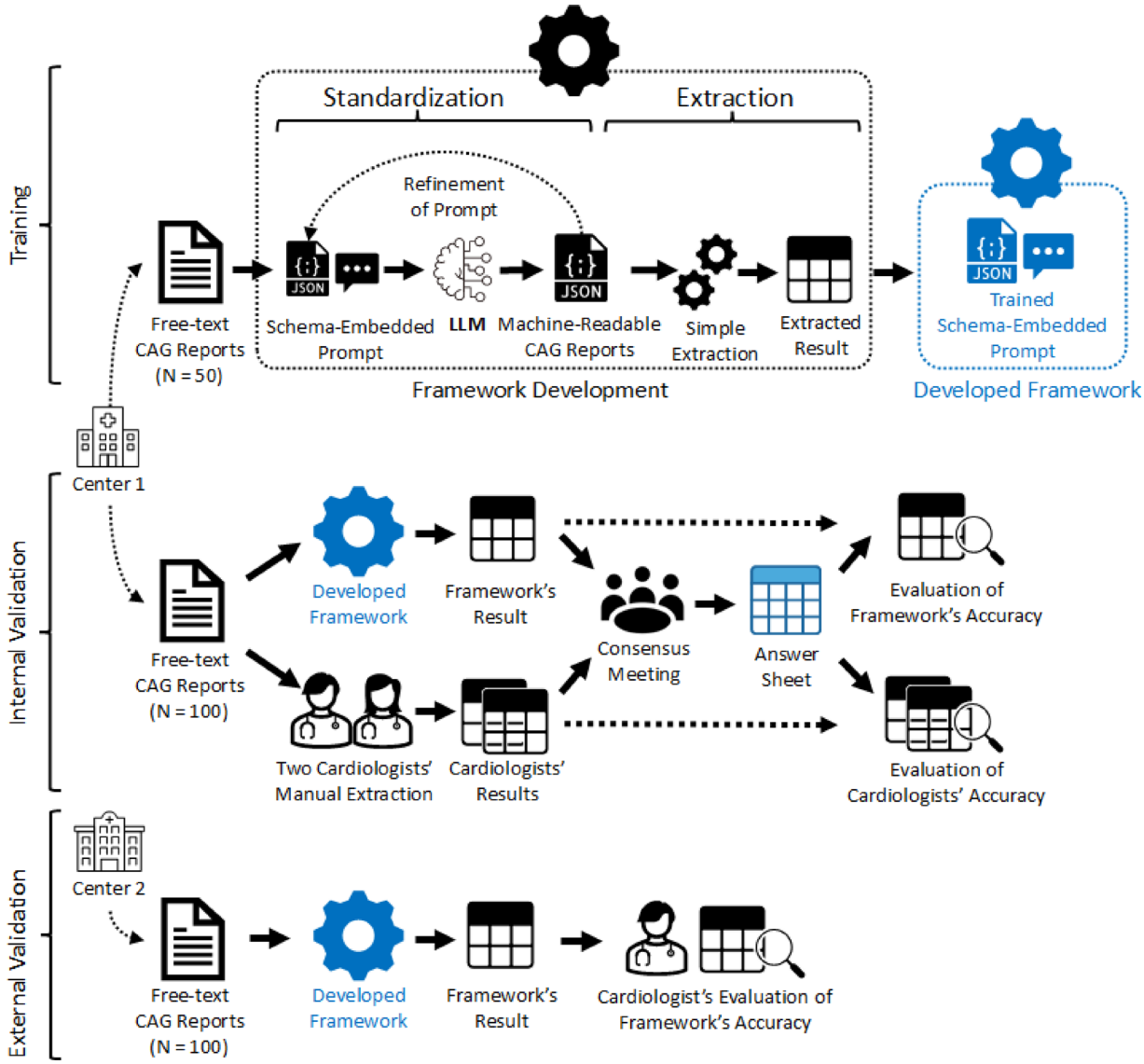

Fig. 3. Overview of the training, internal test, and external test processes of the framework. Fifty CAG reports from Center 1 were used during the training phase to iteratively refine the schema-embedded prompts and extraction logic until stable performance was achieved. For the internal validation, 100 reports from Center 1 were processed by both the developed framework and two cardiologists. A consensus meeting, reviewing both the framework outputs and cardiologists' extractions, produced a gold-standard answer sheet against which accuracy was evaluated. For the external validation, 100 reports from Center 2 were analyzed to assess the framework's generalizability, with cardiologists independently inspecting and confirming each extracted result. Center 1, Severance Hospital; Center 2, Ilsan Hospital. LLM, large language model.

format, which enabled efficient organization of multiple clinical events (e.g., multiple lesions or procedures) and flexible representation of clinical details at various levels of granularity[10–12]. The full annotation schema is shown in Supplementary Fig. 1.

Development of instruction and few-shot prompts

Utilizing the schema, we developed an instruction prompt[21] that consist of designation of a role to LLM as a data scientist, assignment of a task to convert free-text CAG note into JSON format, exhaustive rules to follow, domain-specific knowledge, and the schema to adhere. The domain-specific knowledge included a Synergy Between Percutaneous Coronary Intervention with Taxus and Cardiac Surgery (SYNTAX) score segmentation system[22] for precise documentation of coronary anatomy; lesion morphology and characteristics based on American College of Cardiology/American Heart Association (ACC/AHA) classification[23–25]; and details of the intervention, procedural complications, and outcomes.

Furthermore, we developed a few-shot prompt using eight representative cases from the training dataset that encompassed various clinical scenarios, from simple single-vessel disease to complex multivessel interventions,

ensuring consistency in the interpretation of complex anatomical and procedural details. While these reference cases were utilized for an internal test, they were not applied in the external test, to avoid institutional bias in reporting styles. The full instruction prompt and few-shot prompts are shown in Supplementary Fig. 1 and Supplementary Fig. 2, respectively.

LLM implementation

Building on to the rigorous refinement of the instruction and few-shot prompts, Generative Pretrained Transformer 4 Omni (GPT-4o) from OpenAI[26], one of the commercially-available LLMs, was used as the representative LLM for internal and external test dataset. To test the model-agnostic nature of the proposed framework, we additionally evaluated the same internal test dataset using two alternative large language models, Gemini-2.5-Flash (Google DeepMind)[27] and Claude-4.5-Sonnet (Anthropic)[28]. Each model was implemented in an identical pipeline and prompt structure without further optimization. For practical prompt engineering and validation of the large dataset, a low-code workflow using 'GPT for Excel Word', an extension that is easy to use in Microsoft Excel, was used to directly invoke the API of LLMs inside Excel. Solely with the prompts we developed, anyone who does not know how to code can convert CAG reports into structured format within Excel. A simple implementation example using 'GPT for Excel Word' is provided in Supplementary Method 1 to help readers reproduce the workflow in their own institutions. In accordance with the Minimum Reporting Items for Clear Evaluation of Accuracy Reports of Large Language Models in Healthcare (MINI-CLEAR-LLM) checklist, we disclose the details of this study to ensure transparency in utilizing LLMs for healthcare applications in Supplementary Method 1[29].

Rule-based extraction of 12 key information

Only after the first step of standardization of free-text CAG reports into machine-readable format is completed can the second step of extracting clinically relevant key information be performed through rule-based algorithm.

Two cardiologists (JYJ and HK) pre-defined the following 12 key information points: four CAG-related key points and eight PCI-related key points. The four CAG-related key points included the location of previous stents, previous stent information, the location and type of lesion, and anatomical diagnosis (e.g., two-vessel disease). *Previous stent information* included the device name, diameter, and length of previous stents. *Type of lesion* (A, B1, B2, and C) was determined for each lesion according to the ACC/AHA classification. The eight PCI-related key points focused on the six criteria of complex PCI[30]: Multivessel PCI, implantation of ≥ 3 stents; treatment of ≥ 3 lesions; bifurcation PCI using ≥ 2 stents; total stent length > 60 mm; chronic total occlusion (CTO) as the target lesion[31–35]; complex PCI; and current stent information. *Complex PCI* was defined when any one or more criteria were met in the index PCI. *Current stent information* included the device name, diameter, and length of current stents. The complete extraction algorithm is explained in Supplementary Method 2 and is available at our public repository: https://github.com/jiuisdisciple/CAGtoJSON.

Validation and statistics

After development of the two-step pipeline, the framework's accuracy was assessed through both internal and external tests. In the absence of a task-matched benchmark or public dataset, we selected independent cardiologists' manual extraction as the most clinically meaningful comparator that reflects real-world practice. The primary endpoint was exact-match accuracy at the item level, which is appropriate for this task because each key point has a single, unambiguous target value (e.g., lesion location/type; device name/diameter/length). For the internal test, two experienced cardiologists (JYJ and HSK) independently extracted the 12 key information points manually. We then held a consensus adjudication in which the cardiologists reviewed both their manual extractions and the framework's outputs to produce an item-level gold-standard answer sheet. The framework's performance was compared against the cardiologists' pre-consensus extractions using Fisher's exact test (p value threshold of 0.05). For the external test, cardiologists directly evaluated the framework's extraction results through thorough inspection.

Data availability

The study's underlying data include deidentified free-text CAG notes that contain sensitive clinical information with a risk of re-identification. Therefore they will not be made publicly available. Inquiries should be directed to the designated corresponding author (SCY).

Received: 17 September 2025; Accepted: 8 December 2025
Published online: 03 January 2026

References

1. Rush, E. N. et al. JSONize: A scalable machine learning pipeline to model medical notes as semi-structured documents. *AMIA Jt. Summits Transl Sci. Proc.* **2020**, 533–541 (2020).
2. Hansen, K. N. et al. One-year rehospitalisation after percutaneous coronary intervention: A retrospective analysis. *EuroIntervention* **14**, 926–934. https://doi.org/10.4244/eij-d-17-00800 (2018).
3. Lee, C. et al. Blood pressure and mortality after percutaneous coronary intervention: A population-based cohort study. *Sci. Rep.* **12**, 2768. https://doi.org/10.1038/s41598-022-06627-4 (2022).
4. Waldo, S. W. et al. Temporal trends in coronary angiography and percutaneous coronary intervention. *JACC: Cardiovasc. Interventions*. **11**, 879–888. https://doi.org/10.1016/j.jcin.2018.02.035 (2018).
5. Unlu, O. et al. Retrieval-augmented Generation–Enabled GPT-4 for clinical trial screening. *NEJM AI*. **1**, AIoa2400181. https://doi.org/10.1056/AIoa2400181 (2024).

6. Naidu, S. S. et al. SCAI expert consensus update on best practices in the cardiac catheterization laboratory: This statement was endorsed by the American college of cardiology (ACC), the American heart association (AHA), and the heart rhythm society (HRS) in April 2021. *Catheter Cardiovasc. Interv.* **98**, 255–276 (2021).
7. Park, J. et al. A framework (SOCRATex) for hierarchical annotation of unstructured electronic health records and integration into a standardized medical database: Development and usability study. *JMIR Med. Inf.* **9**, e23983. https://doi.org/10.2196/23983 (2021).
8. OpenAI et al. GPT-4 Technical Report. arXiv:2303.08774 (2023). https://arxiv.org/abs/2303.08774
9. Landolsi, M. Y., Hlaoua, L. & Romdhane, L. B. Extracting and structuring information from the electronic medical text: State of the Art and trendy directions. *Multimedia Tools Appl.* **83**, 21229–21280 (2024).
10. Adams, L. C. et al. Leveraging GPT-4 for post hoc transformation of free-text radiology reports into structured reporting: A multilingual feasibility study. *Radiology* **307**, e230725. https://doi.org/10.1148/radiol.230725 (2023).
11. Huang, J. et al. A critical assessment of using ChatGPT for extracting structured data from clinical notes. *NPJ Digit. Med.* **7**, 106. https://doi.org/10.1038/s41746-024-01079-8 (2024).
12. Cho, H. et al. Extracting lung cancer staging descriptors from pathology reports: A generative Language model approach. *J. Biomed. Inform.* **157**, 104720. https://doi.org/10.1016/j.jbi.2024.104720 (2024).
13. Min, D. et al. Large Language models for CAD-RADS 2.0 extraction from semi-structured coronary CT angiography reports: A multi-institutional study. *Korean J. Radiol.* **26**, 817 (2025).
14. Arnold, P. G. et al. Performance of large Language models for CAD-RADS 2.0 classification derived from cardiac CT reports. *J. Cardiovasc. Comput. Tomogr.* **19**, 322–330. https://doi.org/10.1016/j.jcct.2025.03.007 (2025).
15. Dao, N. et al. Generative artificial intelligence for automated data extraction from unstructured medical text. *JAMIA Open.* **8** https://doi.org/10.1093/jamiaopen/ooaf097 (2025).
16. Barak-Corren, Y. et al. From text to data: automatically extracting data from catheterization reports using generative artificial intelligence. *J. Soc. Cardiovasc. Angiography Interventions*. **4**, 102242 (2025).
17. Turchioe, M. R. et al. Systematic review of current natural Language processing methods and applications in cardiology. *Heart* **108**, 909–916 (2022).
18. You, S. C. et al. Association of Ticagrelor vs clopidogrel with net adverse clinical events in patients with acute coronary syndrome undergoing percutaneous coronary intervention. *Jama* **324**, 1640–1650 (2020).
19. Khan, M. S. et al. Leveraging electronic health records to streamline the conduct of cardiovascular clinical trials. *Eur. Heart J.* **44**, 1890–1909 (2023).
20. Meta *Introducing Meta Llama 3: The Most Capable Openly Available LLM to Date* (2024). https://ai.meta.com/blog/meta-llama-3/
21. Chen, B., Zhang, Z., Langrené, N. & Zhu, S. Unleashing the potential of prompt engineering in large language models: A comprehensive review. arXiv:2310.14735 (2023). https://ui.adsabs.harvard.edu/abs/2023arXiv231014735C
22. Sianos, G. et al. The SYNTAX score: an angiographic tool grading the complexity of coronary artery disease. *EuroIntervention* **1**, 219–227 (2005).
23. Ryan, T. J. Guidelines for percutaneous transluminal coronary angioplasty: A report of the American college of Cardiology/American heart association task force on assessment of diagnostic and therapeutic cardiovascular procedures (Subcommittee on percutaneous transluminal coronary Angioplasty). *J. Am. Coll. Cardiol.* **12**, 529–545. https://doi.org/10.1016/0735-1097(88)90431-7 (1988).
24. Konigstein, M. et al. Utility of the ACC/AHA lesion classification to predict outcomes after contemporary DES treatment: Individual patient data pooled analysis from 7 randomized trials. *J. Am. Heart Association*. **11**, e025275. https://doi.org/10.1161/JAHA.121.025275 (2022).
25. Theuerle, J. et al. Utility of the ACC/AHA lesion classification as a predictor of procedural, 30-day and 12-month outcomes in the contemporary percutaneous coronary intervention era. *Catheter Cardiovasc. Interv*. **92**, E227–e234. https://doi.org/10.1002/ccd.27411 (2018).
26. OpenAI. *Hello GPT-4o* (2024). https://openai.com/index/hello-gpt-4o/
27. Deepmind, G. *Gemini 2.5 Flash, Best for Fast Performance on Everyday Tasks* (2025). https://deepmind.google/models/gemini/flash/
28. Anthropic *Introducing Claude Sonnet 4.5* (2025). https://www.anthropic.com/news/claude-sonnet-4-5
29. Park, S. H., Suh, C. H., Lee, J. H., Kahn, C. E. & Moy, L. Minimum reporting items for clear evaluation of accuracy reports of large language models in healthcare (MI-CLEAR-LLM). *Korean J. Radiol.* **25**, 865–868. https://doi.org/10.3348/kjr.2024.0843 (2024).
30. Giustino, G. et al. Efficacy and safety of dual antiplatelet therapy after complex PCI. *J. Am. Coll. Cardiol.* **68**, 1851–1864. https://doi.org/10.1016/j.jacc.2016.07.760 (2016). https://doi.org/https://doi.org/
31. Suh, J. et al. The relationship and threshold of stent length with regard to risk of stent thrombosis after drug-eluting stent implantation. *JACC Cardiovasc. Interv*. **3**, 383–389 (2010).
32. Mauri, L. et al. Effects of stent length and lesion length on coronary restenosis. *Am. J. Cardiol.* **93**, 1340–1346 (2004).
33. Van Werkum, J. W. et al. Predictors of coronary stent thrombosis: The Dutch stent thrombosis registry. *J. Am. Coll. Cardiol.* **53**, 1399–1409 (2009).
34. Brilakis, E. S. et al. Procedural outcomes of chronic total occlusion percutaneous coronary intervention: A report from the NCDR (National cardiovascular data Registry). *JACC Cardiovasc. Interv*. **8**, 245–253 (2015).
35. Holmes, D. R. et al. Stent thrombosis. *J. Am. Coll. Cardiol.* **56**, 1357–1365 (2010).

Acknowledgements

This research was supported by a grant of the MD-Phd/Medical Scientist Training Program through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea. This research was also supported by the National Institute of Health(NIH) research project(project No.2025ER090500).

Author contributions

JWS: collection and assembly of data, statistical analysis, drafting the manuscriptJYJ: training of the framework, internal and external evaluation, writing and editingHSK: internal test, writing and editingYGK: writing and editingSCY: conception of idea, training of the framework, editing of the manuscriptAll authors revised and approved the final manuscript for submission.

Funding

This research was supported by a grant of the MD-Phd/Medical Scientist Training Program through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea, and also by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number:

HI22C0452).

Declarations

Competing interests

SCY reports grants from Daiichi Sankyo. He is a coinventor of granted Korea Patent DP-2023-1223 and DP-2023-0920, and pending Patent Applications DP-2024-0909, DP-2024-0908, DP-2022-1658, DP-2022-1478, and DP-2022-1365 unrelated to current work. SCY is a chief executive officer of PHI Digital Healthcare. Other authors have no potential conflicts of interest to disclose.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-32150-3.

Correspondence and requests for materials should be addressed to S.C.Y.