机器人及人工智能大赛

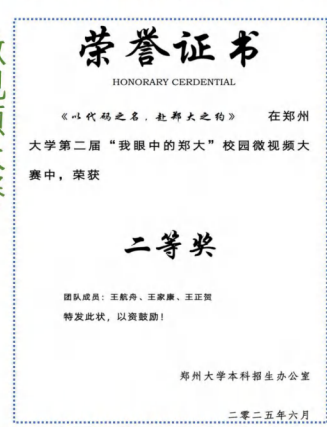
网络技术挑战赛

蓝桥杯

计算机设计大赛

软件著作权(四件申请中)

微视频大赛



河南省三好学生

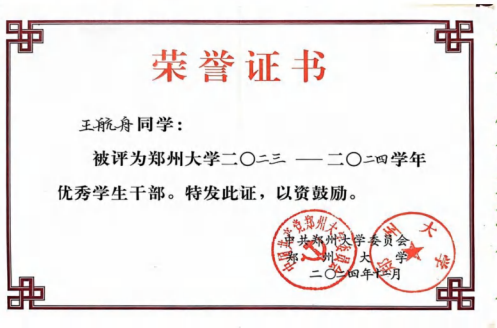
国家励志奖学金

暑期三下乡社会实践

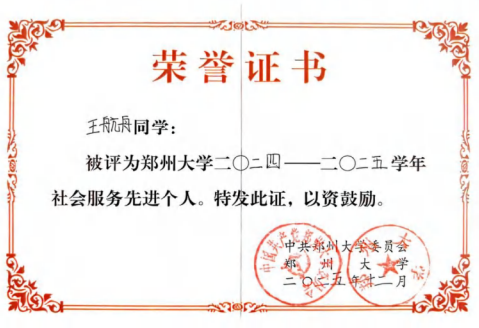
寒假社会实践

三好学生

优秀学生干部



社会服务先进个人



文体活动先进个人

读的仿生解耦式检索增强框架NeuroECG-RAG：面向心电图智能解驱动空间寻址的心电图诊断模型REISA：基于检索增强生成与指令

参与两篇A+医疗方向论文，均已被ICIC 2026录用。

NeuroECG-RAG: A Biomimetic Decoupled Retrieval-Augmented Framework for ECG Interpretation

Wen Li, Ziyu Liu, Hangzhou Wang, Ziyang He, and Yinling Wang(✉)

Zhengzhou University, Zhengzhou 450001, China
brut819@qq.com

Abstract. Electrocardiography (ECG) is essential in clinical diagnosis, but existing ECG interpretation methods still face challenges such as medical hallucinations, limited knowledge updating, and weak evidence-grounded reasoning. To address these issues, we propose NeuroECG-RAG, a biomimetic decoupled retrieval-augmented framework for intelligent ECG interpretation. It follows a perception–retrieval–reasoning pipeline: an ECG visual expert first converts raw ECG images into structured waveform descriptions; then, a biomimetic retrieval module performs multi-hop evidence retrieval over an ECG-oriented knowledge graph built from Wikipedia and The ECG Made Easy; finally, Qwen3.5 generates diagnoses and clinical interpretations under explicit evidence constraints. Experiments on the ECG-Bench core subset show that NeuroECG-RAG achieves the best overall performance, with clear advantages in clinical report generation and multi-turn reasoning while remaining competitive in classification tasks. Overall, this work provides a more reliable and interpretable solution for intelligent ECG analysis.

Keywords: Electrocardiography, Multimodal Large Language Models, Retrieval-Augmented Generation, Biomimetic Memory Retrieval, Clinical Reasoning.

1 Introduction

Electrocardiography (ECG) is a non-invasive and low-cost diagnostic tool widely used in clinical practice, especially for detecting cardiac rhythm abnormalities. Unlike natural images, ECG interpretation depends on fine-grained waveform features, such as P waves, QRS complexes, ST-T changes, rhythm regularity, and cross-lead relationships. Therefore, intelligent ECG interpretation is not only a visual recognition task, but also a knowledge-intensive decision-making process that requires medical criteria, diagnostic thresholds, and clinical reasoning.

Existing ECG interpretation has been extensively studied with deep learning methods, including signal-based and image-based diagnostic models [1,2]. These methods can be broadly divided into two categories. The first category is end-to-end methods, which directly learn the mapping from ECG input to diagnostic output within a single parametric model [3,4]. Recent multimodal large language models (MLLMs) have advanced cross-modal alignment, instruction following, and long-range reasoning, enabling ECG interpretation to move beyond fixed-label prediction toward open-ended tasks such as abnormality question answering, evidence explanation, and report generation. ECG-specialized models, including PULSE [5], MEIT [6], ECG-R1 [7], and GEM [8], further improve waveform understanding by adapting MLLMs to ECG-specific signal and image representations.

However, these methods still rely mainly on parametric knowledge and may suffer from insufficient knowledge integration and medical hallucinations in complex reasoning or knowledge-updating scenarios.

The second category is decoupled methods, which separate waveform perception, evidence retrieval, and language generation into different modules. This design reduces error propagation and improves interpretability, knowledge updating, and hallucination suppression. For example, ECG-Chat [9, 10] connects a frozen vision encoder with a language model through a lightweight bridge, while ViperGPT [10] decouples visual processing from reasoning sequence. In the ECG domain, Q-Heart [2] uses historical clinical report retrieval and dynamic prompting to improve ECG question answering.

However, existing decoupled ECG methods still lack a systematic mechanism for converting visual perception outputs into structured queries and connecting them with external knowledge. Retrieval-augmented generation (RAG) [11,12] provides a useful starting point by introducing external evidence into the generation process to improve factual consistency and reasoning. Several RAG extensions, such as IRCoT [13], Self-RAG [14], RAPTOR [15], Adaptive-RAG [16], and HippoRAG [17], have improved evidence selection and multi-hop reasoning. Nevertheless, most RAG methods are designed for text-query–text-knowledge settings and rely on flat vector retrieval, which is not directly suitable for ECG interpretation, where evidence comes from waveform images and diagnosis often requires multi-hop clinical knowledge integration.

To address this challenge, we propose NeuroECG-RAG, a biomimetic decoupled retrieval-augmented reasoning framework for intelligent ECG interpretation. It follows a clinically inspired workflow: first inspect the waveform, then retrieve diagnostic criteria, and finally generate the diagnosis. Specifically, a visual expert converts ECG images into structured medical descriptions; an open knowledge graph and biomimetic memory retrieval module perform multi-hop evidence retrieval; and Qwen3.5 generates the final diagnosis under the constraints of visual descriptions, retrieved evidence, and task instructions. Compared with existing ECG-RAG methods that mainly rely on handcrafted prompts and flat text retrieval, NeuroECG-RAG integrates ECG-specific visual understanding with structured retrieval-augmented reasoning, supporting more reliable ECG interpretation, report generation, and multi-turn clinical reasoning.

The main contributions of this work are as follows.

1. **A biomimetic decoupled framework for ECG interpretation:** We propose NeuroECG-RAG, which coordinates an ECG visual expert, the Bio-RAG (Artificial Hippocampus), and a reasoning core (Artificial Neocortex) to form a closed-loop pipeline of waveform perception, criterion retrieval, and diagnosis generation.
2. **A large-scale structured ECG knowledge graph for multi-hop evidence retrieval:** We construct a large-scale ECG-oriented open knowledge graph based on Wikipedia and The ECG Made Easy [18], containing phrase nodes, passage nodes, relation edges, synonym edges, and contextual edges, thereby supporting multi-hop associative retrieval and context-preserving evidence grounding.
3. **A retrieval-augmented reasoning pipeline for grounded diagnosis generation:** We integrate structured visual descriptions with graph-based retrieval and evidence-constrained generation, allowing the model to produce diagnoses and explanations that are more explicitly supported by external clinical knowledge.
4. **Comprehensive evaluation on ECG-Bench core tasks:** Experimental results on the core subset of ECG-Bench show that NeuroECG-RAG achieves strong overall performance, with particular advantages in report generation and multi-turn clinical reasoning, while maintaining competitive classification capability.

Fig. 1. Framework of NeuroECG-RAG.

2 Method

2.1 Structured Textual Description of ECG Images via the Visual Expert.

As shown in Fig. 1, the Visual Expert is built upon ECGInstruct-driven multimodal instruction tuning [5], covering waveform recognition, rhythm analysis, morphological assessment, and report generation. Since this paradigm has shown stronger fine-grained ECG perception and cross-domain generalization than general-purpose multimodal models on ECG-Bench, we adopt it as the front-end perception module.

Instead of reconstructing ECG images into one-dimensional signals before text generation [19], we directly use an ECG-pretrained multimodal model to generate structured natural language descriptions from raw ECG images. This avoids the amplification of grid artifacts, sampling distortion, and blurring noise often introduced during signal reconstruction. It also makes the generated descriptions more semantically aligned with triples and entity relations in the knowledge graph.

For each ECG image, the Visual Expert produces a structured description covering rhythm regularity, P-wave visibility and morphology, PR interval, QRS abnormalities, ST–T changes across leads, and potential pacing or conduction-related findings. These observations serve as objective waveform evidence and are used to form the joint query for subsequent retrieval and reasoning.

2.2 ECG Knowledge Corpus and Knowledge Graph Construction

We construct an offline indexing mechanism that integrates an open knowledge graph, passage-level representations, and online LLM-based filtering. This structure organizes ECG knowledge while preserving both conceptual relations and contextual information.

The knowledge sources include Wikipedia entries related to ECG diagnostic labels, arrhythmias, conduction blocks, and ST–T abnormalities, as well as diagnostic descriptions from The ECG Made Easy [18]. They cover rhythm identification, waveform morphology, lead distribution, and diagnostic criteria. In total, the corpus contains 4,043 passages.

During offline indexing, the corpus is first segmented into semantically complete passages. GPT-4o-mini is then used for Open Information Extraction (OpenIE), producing schema-free triples. The subjects and objects of triples are treated as phrase nodes, predicates as relation edges, and the original passages as passage nodes. This enables joint modeling of conceptual knowledge and source context.

To improve connectivity among different ECG expressions, we compute embedding similarities between phrase nodes using NV-Embed-v2 and add synonym edges. For example, "absence of P wave," "no visible P wave," and "P wave not discernible" refer to the same diagnostic cue, while "QRS widening," "prolonged QRS duration," and "QRS $\geq$ 120 ms" may indicate the same abnormal pattern.

The final knowledge graph $\mathcal{G}$ contains phrase nodes, passage nodes, relation edges, synonym edges, and contextual edges. It supports multi-hop associative reasoning along the chain of concept $\rightarrow$ feature $\rightarrow$ diagnostic criterion $\rightarrow$ source passage.

2.3 Biomimetic Memory-Driven Online Associative Retrieval

In the online retrieval stage, the user query is concatenated with the structured description generated by the Visual Expert to form a joint query q^*. This integrates visual observations with task intent. The retrieval process includes five steps: query-to-triple retrieval, recognition memory, seed assignment, Personalized PageRank propagation, and evidence ranking.

First, NV-Embed-v2 computes the semantic similarity between the joint query and triples in the graph, returning candidate triples as the starting points for graph propagation:

$$s_t = \cos\left(\mathbf{M}(\mathbf{q}^*), \mathbf{M}(t)\right), \quad t \in \mathcal{T} \tag{1}$$

$$\mathcal{T}_{\text{init}} = \text{Top}\mathbf{M}_{\text{ret}}\, s_t \tag{2}$$

where $\mathbf{M}(\cdot)$ denotes the embedding encoder, $t \in \mathcal{T}$represents candidate triples in the knowledge graph, $s_t$is the semantic similarity between the joint query and triple t, and $\mathcal{T}$opM ($\cdot$)returns the top $\mathcal{M}$triples.

Next, GPT-4o-mini filters the candidate triples and retains only those truly relevant to the current query. This step reduces semantic noise from dense retrieval, which is important for ECG interpretation because one feature, such as QRS widening, may correspond to multiple diagnostic directions, including bundle branch block, ventricular rhythm, or pacing.

$$\mathcal{T}^- = \{t \in \mathcal{T}_{\text{init}} \mid \phi(\mathbf{q}^*, t) = \text{"irrelevant"}\} \tag{4}$$

where $\phi(t \mid q^*)$denotes the estimated relevance probability of triple t, $p_\phi(\cdot)$is the filtering model, $r_t$is the relevance threshold, and $\mathcal{T}^-$is the filtered triple set.

After filtering, the remaining triples are assigned to phrase nodes and passage nodes. The passage node weight scaling factor is set to 0.05, balancing concept-level propagation and contextual preservation. Phrase node weights are computed as follows:

$$\mathcal{T}_v = \{\mathbf{t} \mid \mathbf{t} \in \mathcal{T}, \mathbf{v} \in \mathbf{t}\} \tag{5}$$

$$w_v = \frac{1}{|\mathcal{T}_v|} \sum_{t \in \mathcal{T}_v} s_t \tag{6}$$

Here, $\mathcal{T}_v$denotes the filtered triples containing phrase node v, $s_t$is the node-specific weight, and $w_v$is the initial seed score of the phrase node.

To preserve contextual information from the original passages, we further assign weighted seed scores to passage nodes and construct a personalized restart distribution accordingly:

$$w_p = \beta \cdot \cos\left(\mathbf{M}(\mathbf{q}^*), \mathbf{M}(p)\right), \quad p \in \mathcal{V}_{\text{psg}} \tag{7}$$

$$r_v = \frac{w_v}{\sum_{u \in \mathcal{V}_{\text{ent}} \cup \mathcal{V}_{\text{psg}}} w_u} \tag{8}$$

where $p \in \mathcal{V}_{\text{psg}}$denotes a passage node, $\beta$is the scaling factor for passage node weights, $\mathcal{V}_{\text{psg}}$is the set of all seed nodes, and $r_v$is the normalized restart probability of node v.

The system then performs Personalized PageRank propagation over the graph:

$$\pi = \alpha \mathbf{W}^\top \pi + (1-\alpha)\mathbf{r} \tag{9}$$

where $\mathbf{W}$is the transition matrix, $\pi$is the personalized restart distribution, $\alpha = 0.5$is the damping factor, and $\mathbf{r}$is the initial restart distribution. Compared with flat vector retrieval, PPR enables multi-hop propagation along graph edges and activates evidence that may not be directly matched but is clinically relevant. Finally, passages are ranked by propagated scores, and the top-k passages are selected as the final evidence set $\mathcal{E}$.

2.4 Evidence-Constrained Diagnosis Generation with Qwen3.5

In the final stage, Qwen3.5-27B is used as the text generation backbone for diagnosis generation and clinical interpretation under explicit evidence constraints. We select this model for its strong long-context modeling capability [20], which allows it to integrate structured waveform descriptions, retrieved evidence passages, and task instructions in retrieval-augmented settings.

Since this stage no longer requires direct image processing, Qwen3.5-27B is deployed in a text-only manner. This design avoids reintroducing a generic visual pathway and focuses computation on reasoning over retrieval-augmented textual inputs.

The input to Qwen3.5-27B consists of system instructions, structured waveform descriptions, retrieved knowledge passages, and the user query. The system prompt requires the model to prioritize the visual expert's descriptions and the retrieved evidence. If the evidence is insufficient, contradictory, or unable to support a confident conclusion, the model is instructed to provide a conservative response rather than unsupported speculation.

Through this design, Qwen3.5-27B generates diagnoses within the semantic boundary defined by visual observations and retrieved evidence. As shown in Fig. 2, the output is no longer unconstrained open-ended generation, but evidence-grounded diagnosis generation, reducing the risk of hallucination.

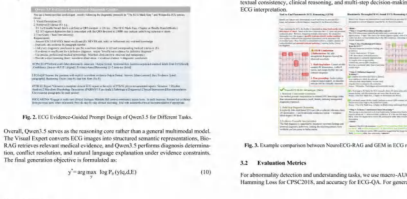


Fig. 2. ECG Evidence-Guided Prompt Design of Qwen3.5 for Different Tasks.

Overall, Qwen3.5 serves as the reasoning core rather than a general multimodal model. The Visual Expert converts ECG images into structured semantic representations, Bio-RAG retrieves supporting clinical evidence, and Qwen3.5 performs diagnosis determination, conflict resolution, and natural language explanation under evidence constraints. The final generation objective is formulated as

$$\mathcal{Y}^* = \arg\max_{\mathcal{Y}} \log P_\theta\left(\mathcal{Y} \mid q, \mathcal{E}\right) \tag{10}$$

where $P_\theta(\cdot)$ denotes the conditional generation probability of Qwen3.5, $\mathcal{Y}$ denotes a candidate answer, and q^* represents the optimal output conditioned on the visual description $\mathbf{d}$, the retrieved evidence $\mathcal{E}$, and the user query q.

3 Experiments

3.1 Benchmark Datasets

To reduce computational cost and avoid redundant evaluation bias from using the full ECGBench, we evaluate NeuroECG-RAG on a compact core subset covering four representative ECG tasks. The subset includes CPSC2018, ECG-QA, PTB-XL Report, and ECG Arena, with 2,061, 1,317, 500, and 50 samples, respectively.

CPSC2018 is used for multi-label abnormality detection across nine diagnostic categories and evaluates cross-domain diagnostic generalization. ECG-QA contains ECG image–based, question-answer pairs over seven question templates, assessing basic ECG understanding and abnormality recognition. PTB-XL Report includes 500 physician-annotated reports and evaluates long-form clinical report generation, focusing on diagnostic accuracy, waveform description, and clinical completeness. ECG Arena consists of 50 manually curated multi-turn clinical dialogues, designed to assess contextual consistency, clinical reasoning, and multi-step decision-making in open-ended ECG interpretation.

Fig. 3. Example comparison between NeuroECG-RAG and GEM in ECG report generation.

3.2 Evaluation Metrics

For abnormality detection and understanding tasks, we use macro-AUC, macro-F1, and Hamming Loss for CPSC2018, and accuracy for ECG-QA. For generation tasks, GPT-

14. Asai, A., Wu, Z., Wang, Y., Sil, A., Hajishirzi, H.: Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection. In: International Conference on Learning Representations (ICLR) (2024)
15. Sarthi, P., Abdullah, S., Tuli, A., Khanna, S., Goldie, A., Manning, C.D.: RAPTOR: Recursive Abstractive Processing for Tree-Organized Retrieval. In: International Conference on Learning Representations (ICLR) (2024)
16. Jeong, S., Baek, J., Cho, S., Hwang, S.J., Park, J.C.: Adaptive-RAG: Learning to Adapt Retrieval-Augmented Large Language Models through Question Complexity. In: Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), pp. 7036–7050 (2024)
17. Gutiérrez, B.J., Shu, Y., Gu, Y., Yasunaga, M., Su, Y.: HippoRAG: Neurobiologically Inspired Long-Term Memory for Large Language Models. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 59532–59569 (2024)
18. Wampler, J., Hampton, A., Adlam, D.: The ECG Made Easy. 10th edn. Elsevier (2024)
19. Liu, C., Wan, Z., Wang, H., et al.: Large Language Models Are Few-Shot Health Learners. arXiv preprint arXiv:2305.15525 (2023)
20. Yang, S., Kautz, J., Hatamizadeh, A.: Gated Delta Networks: Improving Mamba2 with Delta Rule. arXiv preprint arXiv:2412.06464 (2024)

4o is used to evaluate PTB-XL Report outputs against physician-annotated references from three aspects: diagnostic accuracy, waveform description, and rhythm analysis. Each aspect is scored on a 0–10 scale and normalized to [0,100]. For ECG Arena, GPT-4o evaluates multi-turn responses in terms of accuracy, completeness, and instruction adherence, with the final score also scaled to 100.

3.3 Experimental Settings

Experiments were conducted on a server equipped with a server equipped with 8 NVIDIA GeForce RTX 4090 GPUs. We use a quantized version of Qwen3.5-27B [5] to generate structured ECG descriptions; during the graph retrieval stage, the GPT-4o-mini API is used for relation filtering. The vector retriever is BGE-small-en-v1.5 used for dense retrieval and similarity computation; the reasoning stage adopts Qwen3.5-27B as the text generation backbone.

In the offline graph construction stage, the synonym edge threshold is set to 0.8; the damping factor of personalized PageRank is set to 0.5; the passage node weight scaling factor is set to 0.05; and 5 phrase nodes are retained as seed nodes during graph propagation initialization.

For decoding strategies, CPSC2018 and ECG-QA adopt the non-thinking/instruct mode with parameters set to temperature = 0.7, top_p = 0.8, and top_k = 20. PTB-XL Report and ECG Arena adopt the thinking mode with parameters set to temperature = 1.0, top_p = 0.95, and top_k = 20. All other parameters remain at their default settings.

3.4 Experimental Results

Table 1. Experimental Results on Four Core Tasks.

Category	Datasets	CPSC 2018		ECG-QA	PTB-XL Report	ECG Arena
	Metric	AUC	F1	HL	Accuracy	Report Score	Arena Score
Proprietary MLLMs	Random	51.2	15.1	28.8	16.2	0	0
	GPT-4o	50.9	10.6	18.2	35.2	50.2	33.5
	GPT-5-mini	49.2	11.0	23.5	14.9	57.1	30.1
	Claude 3.5 Sonnet	52.8	11.5	18.9	34.2	43.7	37.1
Open-source MLLMs	LLaVA-Next-Vicuna-7B	51.5	12.9	20.4	25.0	40.6	15.5
	Deepseek-VL-Chat-7B	50.7	6.0	20.0	21.1	15.6	15.3
	Mantis-8B-siglip-Llama3	51.3	19.1	48.5	23.8	16.6	13.6
ECG-specialized MLLMs	PULSE	74.9	37.6	8.6	**73.8**	61.3	39.9
	GEM	79.1	61.3	8.3	73.6	67.1	49.0
Ours	NeuroECG-RAG	**79.2**	**59.8**	8.9	73.4	**67.9**	**42.0**

To evaluate the contribution of each core component, we design five ablation variants, as shown in Table 2. Removing the visual expert and using only Bio-RAG with Qwen3.5 results in a weak text-only baseline, since the system can only retrieve evidence from the user query and cannot reliably activate ECG-specific diagnostic nodes.

The visual Expert + Bio-RAG variant achieves a substantial gain over the visual expert itself. It performs well on basic abnormality recognition, achieving 57.6, 73.8, 61.3, and 28.9 on CPSC2018, ECG-QA, PTB-XL Report, and ECG Arena, respectively, but remains limited in report generation and multi-hop reasoning due to the lack of full evidence integration.

Adding Qwen3.5 to the visual expert improves PTB-XL Report and ECG Arena scores to 62.2 and 39.0, suggesting that Qwen3.5 helps organize long-form explanations. However, performance drops on CPSC2018 and ECG-QA, indicating that additional abstraction may lose fine-grained threshold-sensitive ECG details.

When standard flat vector retrieval is added, the model further improves on PTB-XL Report and ECG Arena, reaching 63.4 and 41.1, but decreases on CPSC2018 and changes little on ECG-QA. This suggests that standard RAG provides useful textual evidence for generation and reasoning, but may introduce noise for label prediction tasks that depend on precise waveform cues.

The full model, Visual Expert + Bio-RAG + Qwen3.5, achieves the most balanced performance, with scores of 59.8, 73.4, 67.5, and 42.8 across the four tasks. Compared with standard RAG, its gains come from structured association and multi-hop retrieval rather than retrieval alone: the visual expert provides decoded ECG attributes, Bio-RAG organizes evidence chains through triple filtering and Personalized PageRank, and Qwen3.5 generates diagnoses under explicit evidence constraints.

Table 2. Ablation Study Results.

Class	CPSC 2018	ECG-QA	PTB-XL Report	ECG Arena
Bio-RAG+Qwen3.5	59.0	64.22	0	0
Visual Expert+Bio-RAG	57.6	73.8	61.3	38.9
Visual Expert+Qwen3.5	57.2	72.8	62.2	39.0
Visual Expert+RAG+Qwen3.5	56.8	73.1	63.4	41.1
Visual Expert+Bio-RAG+Qwen3.5	**59.8**	**73.4**	**67.5**	**42.8**

3.5 Ablation Experiments

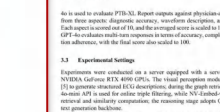


Fig. 4. Comparison of Key Capability Dimensions among NeuroECG-RAG, GEM, and PULSE.

4 Conclusion

In this study, we propose NeuroECG-RAG, a biomimetic decoupled retrieval-augmented framework for intelligent ECG interpretation. It integrates ECG understanding into structured perception, knowledge retrieval, and evidence-constrained reasoning. A visual expert first extracts fine-grained waveform features from ECG images, which are then linked to an open knowledge graph for multi-hop evidence retrieval. The retrieved evidence is finally used to generate diagnoses under explicit constraints. Experiments show that NeuroECG-RAG performs well in report generation, multi-turn question answering, and complex clinical reasoning, while remaining competitive in classification tasks. Overall, this work offers a more reliable and interpretable solution for intelligent ECG interpretation.

1. Petersen, G., Ståhlspetsen, I., Kilsintis, V., Tzavelis, A., Rogers, J.A., Katsaggelos, A.K., Magleaves, N.: State-of-the-Art Deep Learning Methods on Electrocardiogram Data: Systematic Review. JMIR Medical Informatics 10(8), e34474 (2022)
2. Sangha, V., Mortazavi, B.J., Haimovich, A.D., Ribeiro, A.H., Brandt, C.A., Jacoby, D.L., Schulz, W.L., Krumholz, H.M., Ribeiro, A.L.P., Khera, R.: Automated Multilabel Diagnosis on Electrocardiographic Images and Signals. Nature Communications 13(1), 1583 (2022)
3. Pham, H.M., Tang, J., Saeed, A., Ma, D. Q-Heart: ECG Question Answering via Knowledge-Informed Multimodal LLMs. arXiv preprint arXiv:2505.06296 (2025)
4. Ribeiro, A.H., Ribeiro, M.H., Paixão, G.M.M., Oliveira, D.M., Gomes, P.R., Canazart, J.A., Ferreira, M.P.S., Andersson, C.R., MacFarlane, P.W., Meira Jr, W., et al.: Automatic Diagnosis of the 12-Lead ECG Using a Deep Neural Network. Nature Communications 11(1), 1760 (2020)
5. Liu, H., Bai, W., Yao, X., Zhang, Z.: Pioneering Multimodal LLMs to Comprehend 12-Lead Electrocardiographic Images and Digital Medicine (2026)
6. Wan, Z., Liu, Q., Wang, X., Tan, C., Shen, H., Xiong, J., Arcucci, R., Yao, H., Zhang, M.: MEIT: Multimodal Electrocardiogram Instruction Tuning on Large Language Models for Report Generation. In: Findings of the Association for Computational Linguistics: ACL (Findings of ACL), pp. 14519–14527 (2025)
7. Jin, J., Wang, H., Wang, X., Tan, C., Shen, H., Xiong, J., Arcucci, R., Yao, H., Zhang, M.: MEIT: Multimodal Electrocardiogram Instruction Tuning on Large Language Models for Report Generation. In: Findings of the Association for Computational Linguistics: ACL (Findings of ACL), pp. 14519–14527 (2025)
8. Lan, X., Wu, F., He, K., Zhao, Q., Hong, S., Feng, M.: GEM: Empowering MLLM for Grounded ECG Understanding with Time Series and Images. arXiv preprint arXiv:2503.06073 (2025)
9. Li, J., Li, D., Savarese, S., Hoi, S.: BLIP-2: Bootstrapping Language-Image Pre-Training with Frozen Image Encoders and Large Language Models. In: International Conference on Machine Learning (ICML), pp. 19730–19742 (2023)
10. Surís, D., Menon, S., Vondrick, C.: ViperGPT: Visual Inference via Python Execution for Reasoning. In: IEEE/CVF International Conference on Computer Vision (ICCV), pp. 11888–11898 (2023)
11. Lewis, P., Lee, K., Tung, Z., Pasupat, P., Chang, M.: Retrieval-Augmented Language Model Pre-Training. In: International Conference on Machine Learning (ICML), pp. 3929–3938 (2020)
12. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., et al.: Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 9459–9474 (2020)
13. Trivedi, H., Balasubramanian, N., Khot, T., Sabharwal, A.: Interleaving Retrieval with Chain-of-Thought Reasoning for Knowledge-Intensive Multi-Step Questions. In: Annual Meeting of the Association for Computational Linguistics (ACL), pp. 10014–10037 (2023)

REISA: An Electrocardiogram Diagnostic Model Based on Retrieval-Augmented Generation and Instruction-Driven Spatial Addressing

Mingrui Li[1]0009-0008-7087-6439], Yinling Wang[1*], Hangzhou Wang[1], Ziyang He[1]

[1] School of Cyberspace Security, Zhengzhou University, Zhengzhou, China
yinlingwang@zzu.edu.cn

Abstract. Current ECG diagnostic models effectively identify cardiac pathologies in open domains but struggle to establish granular semantic-morphological rela-tionships. Existing paradigms rely heavily on Large Language Models for di-agnostic synthesis, which carries inherent risks of medical hallucinations, and often lack spatial awareness in multi-lead signal processing for location-sensitive diseases. To address these challenges, we propose REISA, an evi-dence-based ECG diagnostic framework with enhanced spatial perception. REISA integrates Retrieval-Augmented Generation (RAG) to fuse original clinical reports with authoritative morphological criteria, thereby anchoring di-agnostic logic to clinical evidence rather than the internal parametric memory of LLMs. Furthermore, we introduce a Morphological Semantic Dehydration Strategy (MSDS) to recon-struct hybrid reports into high-signal-to-noise ratio morphological descriptors by filtering out non-pathological narrative noise via strict physical constraint proto-cols. Central to our framework is the Instruction-driven Spatial Addressing (ISSA) mechanism, which utilizes the instruction-following capabilities of text encoders to transform descriptors into dynamic spatial query vectors, enabling precise addressing of lead-specific regions with-in signal feature maps. System-atic cross-domain zero-shot evaluations across multiple clinical datasets demon-strate that REISA effectively corrects the logi-cal biases of conventional models, achieving substantial improvements in di-agnostic accuracy and robustness. Ul-timately, REISA provides a robust ap-proach for constructing evidence-based ECG diagnostic systems, demonstrat-ing significant clinical utility for automated ECG diagnosis in diverse practical environments.

Keywords: Electrocardiogram, Retrieval-augmented generation, Morphological semantic dehydration, Instruction-driven spatial addressing.

1 Introduction

1.1 A Subsection Sample

Cardiovascular diseases (CVDs) are the leading cause of death globally and a primary cause of physical disability and reduced quality of life , with their incidence showing a continuous upward trend [1]. As a vital standard for clinical cardiac assessment and early diagnostic screening, electrocardiogram (ECG) interpretation depends heavily on professional expertise and is susceptible to inter-observer variability [2]. Furthermore, in remote or medical resource-deficient areas, the scarcity of professional cardiologists often leads to delays in diagnosis. These challenges underscore the need for robust, automated ECG interpretation systems.

Driven by the advancement of deep learning, data-driven automated methods have achieved significant breakthroughs in tasks such as arrhythmia recognition[3]. Early studies utilized 1-D convolutional neural networks (1D-CNN), recurrent neural networks (RNN), and their variants to improve classification accuracy, or incorporated entropy features and attention mechanisms to enhance cross-temporal modeling[4]. To address the issue of high annotation costs, self-supervised learning strategies have also been employed to optimize representations of unlabeled ECG signals[5]. Nevertheless, traditional deep learning still faces significant limitations: first, these models lack an understanding of medical knowledge, diagnostic logic, and clinical semantics, resulting in weak interpretability and a clear divergence from clinical reasoning; second, most models focus exclusively on signal waveform learning and struggle to integrate textual information such as medical history and symptoms, which limits their comprehensive judgment capability for complex ECG abnormalities.

In recent years, the rise of Large Language Models (LLMs) has propelled a transformation in ECG diagnosis from simple pattern recognition toward knowledge-integrated intelligent reasoning, aligning diagnostic processes more closely with the interpretation logic of clinicians [6]. Nevertheless, LLMs face significant limitations. On one hand, due to the lack of specialized medical knowledge, general-purpose LLMs that rely heavily on general corpora are prone to medical hallucinations, leading to label distortion and logical contradictions in diagnostic results, which makes it difficult to meet the stringent requirements of actual diagnostic scenarios [7]. On the other hand, since ECG signals possess inherent temporal continuity, subtle yet critical pathophysiological changes are typically anchored to specific lead channels—traditional encoders lack instruction-following capabilities. Such data-entry and query errors can severely mislead subordinate within feature maps through the content of text labels, which may result in excessively low diagnostic accuracy for location-sensitive diseases [8]. To address these issues, researchers have developed domain-specific ECG diagnostic models, achieving significant breakthroughs in cross-domain and clinical semantics through large-scale ECG-text paired pre-training. For instance, Yu et al. implemented zero-shot diagnosis by transforming ECG features into prompts, breaking through the limitations of small-sample scenarios [9]. The CVDLLM framework proposed by Qiu et al. enhances reasoning capabilities and interpretability by fusing semantic embeddings with graph attention features [10]. Furthermore, some studies have begun to introduce Retrieval-Augmented Generation (RAG) methods, providing models with professional knowledge to mitigate the inherent structural deficiencies of LLMs in this field [11].

To address these challenges, this manuscript proposes REISA, an ECG zero-shot diagnostic model based on Retrieval-Augmented Generation (RAG). This model utilizes RAG technology to generate augmented diagnostic reports, introduces a "morphological bottleneck" strategy to filter textual noise, and employs the ISSA mechanism to achieve targeted addressing and feature extraction within 12-lead spatio-temporal feature maps. Our main contributions are summarized as follows:

1. We utilize MedCPT to retrieve information from a knowledge base constructed based on AHA/ACC guidelines [12]. By performing semantic fusion to reconstruct morphological descriptors with clinical certainty, we alleviate the medical hallucination caused by the over-reliance of LLMs on internal parametric knowledge.
2. By formulating morphological constraint protocols, we reconstruct original reports and evidence-based evidence into high-signal-to-noise ratio (SNR) descriptors, significantly reducing textual redundancy and enhancing the capture precision of the entropy features and attention mechanisms to enhance cross-temporal modeling[8]. To address the issue of high annotation costs, self-supervised learning strategies have also been employed to optimize representations of unlabeled ECG signals[5]. Nevertheless, traditional deep learning still faces significant limitations: first, these models lack an understanding of medical knowledge, diagnostic logic, and clinical semantics, resulting in weak interpretability and a clear divergence from clinical reasoning; second, most models focus exclusively on signal waveform learning and struggle to integrate textual information such as medical history and symptoms, which limits their comprehensive judgment capability for complex ECG abnormalities.
3. We introduce the ISSA mechanism, which leverages the instruction-following capabilities of text encoders to transform descriptors into dynamic spatial query vectors, enabling precise addressing of lead-specific regions with-in signal feature maps. This mechanism effectively corrects the logi-cal biases of conventional models, achieving substantial improvements in di-agnostic accuracy and robustness.
4. Experimental Results

We compared the zero-shot performance of REISA with existing baseline models, including the fully supervised model vResNet1d [16], the contrastive learning-based CLOCS [17], and multimodal models such as Table 2.

Table 2. Comparative analysis of zero-shot diagnostic performance.

Dataset	Method	Learning Paradigm	Knowledge Paradigm	AUROC	F1-score	MCC
PTB-XL	CLOCS	SSL	N/A	0.4980	0.4210	0.3338
	KED	Fronzlo-tion	Generative Narrative	0.7681	0.4800	-
	ESE	Fronzlo-tion	Augmented Context	0.7370	0.5210	-
	REISA(Ours)	Fronzlo-tion	Evidence-Anchored Instruction	0.7926	0.5703	0.2664
CPSC2018	CLOCS	SSL	N/A	0.7120	0.4380	0.3020
	KED	Fronzlo-tion	Generative Narrative	0.8795	0.5120	-
CVDLLM	Fronzlo-tion	Case-based Graph		0.7020	0.4950	-
REISA(Ours)	Fronzlo-tion	Evidence-Anchored Instruction		0.8232	0.4375	0.4105

Experimental results are shown in Table 2, where REISA demonstrates significant performance advantages in zero-shot diagnostic tasks. On the PTB-XL dataset, it achieved an average AUROC of 0.7926, not only outperforming the signal self-supervised model CLOCS but also surpassing existing multimodal foundation models KED and ESE. On the CPSC2018 dataset, REISA reached an AUROC of 0.8232 and an F1-score of 0.4375, further validating the effectiveness of its knowledge-driven strategy in cross-domain generalization. Notably, REISA maintains high diagnostic performance without relying on task-specific fine-tuning, which can be attributed to its "morphological bottleneck" strategy. By filtering out non-pathological narrative noise via strict physical constraint protocols, REISA significantly reduces the logical biases of conventional models, paving the way for more reliable and interpretable solutions for intelligent ECG diagnosis in diverse practical environments.

Ablation Experiments

To validate the contribution of each component in REISA, we designed five ablation variants: (1) removing the RAG module and relying solely on the LLM's internal knowledge; (2) removing the MSDS strategy and directly using raw clinical reports as input; (3) removing the ISSA mechanism and replacing it with standard global average pooling; (4) removing the Instruction-driven Spatial Addressing (ISSA) module in Section 3.4.

This comparison quantifies the contribution of each component to the overall performance. As shown in Table 3, REISA demonstrates significant performance advantages in zero-shot diagnostic tasks. On the PTB-XL dataset, it achieved an average AUROC of 0.7926, not only outperforming the signal self-supervised model CLOCS but also surpassing existing multimodal foundation models KED and ESE. On the CPSC2018 dataset, REISA reached an AUROC of 0.8232 and an F1-score of 0.4375, further validating the effectiveness of its knowledge-driven strategy in cross-domain generalization. Notably, REISA maintains high diagnostic performance without relying on task-specific fine-tuning, which can be attributed to its "morphological bottleneck" strategy. By filtering out non-pathological narrative noise via strict physical constraint protocols, REISA significantly reduces the logical biases of conventional models, achieving substantial improvements in di-agnostic accuracy and robustness. Ultimately, REISA provides a robust ap-proach for constructing evidence-based ECG diagnostic systems, demonstrat-ing significant clinical utility for automated ECG diagnosis in diverse practical environments.

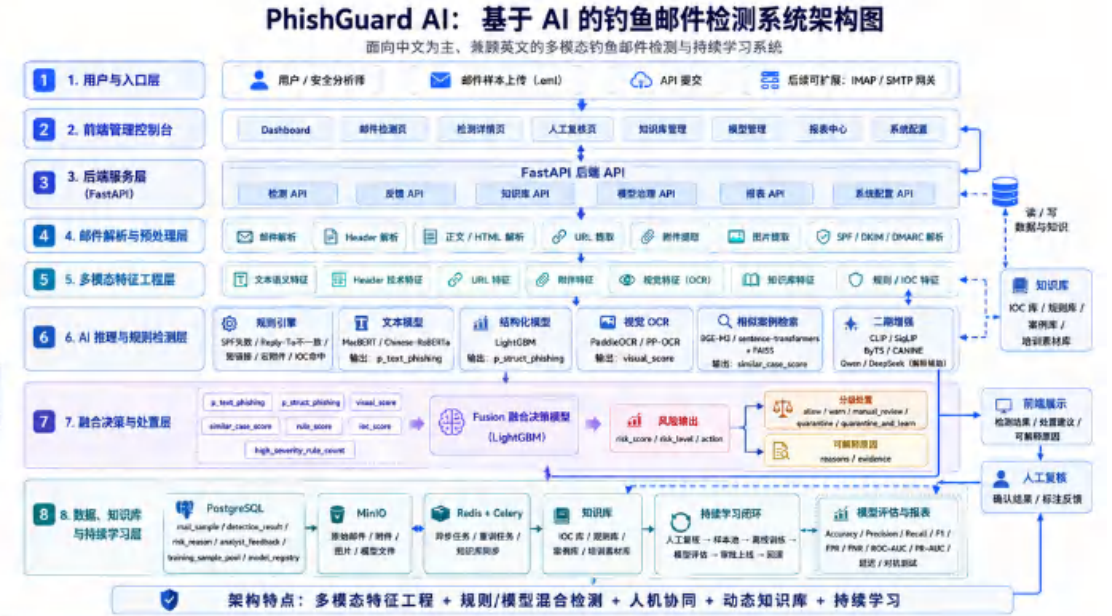
Table 3.

链隐协同：基于区块链和深度隐写的多通道隐蔽信息传输系统
1 用户接入层
Web 可视化大屏 (React/Vite)
微信小程序
PyQt 桌面端 / 静态 Dashboard
2 应用服务层
Flask Web 服务网关
REST API
WebSocket 实时状态推送
用户登录与身份属性管理
文本消息发送 / 接收
图像隐写发送 / 提取
历史记录与审计查询
链上信息与节点监控
3 核心安全能力层
SM2 / SM3 / SM4 国密算法
CP-ABE 属性基加密
SM2 环签名 身份匿名
Shamir 秘密共享分片
LSB 图像隐写
SNR 深度学习图像隐写
AI 载体图像生成
模拟社交信道零宽字符分片
贯穿全流程
4 基础设施层
FISCO BCOS 联盟链
链上存证 交易广播 区块扫描
IPFS 分布式存储
密文文件 模型图像 索引包 CID
MySQL 数据库
用户 密钥 历史记录 社交分析
SNR PyTorch 微服务
图像编码与解码
核心数据流
明文消息 / Secret 图像
SM2 或 CP-ABE 加密
上传 IPFS 获取 CID
LSB / SNR 图像隐写
对载密图 CID 做环签名与 Shamir 分片
一并写入 FISCO BCOS 区块链
一并写入 模拟社交信道
接收端从链上与社交信道重构 CID
从 IPFS 下载载体
提取 CID / 数据 / 恢复明文或秘密图像
5 安全治理与审计层
DLP 敏感数据识别
零信任 动态授权
策略推荐与风险确认
审计证据链
安全态势统计
SM2 + ML-KEM + SM4 后量子混合封装
数据安全
隐蔽通信
链上审计
多通道分片恢复

CroX-Seek 全天候跨模态行人重识别系统
Transformer ReID + RGB-IR 跨模态检索 + 视频目标搜索 + 特征库云边协同
输入层
1. 目标人物图片
2. 监控视频流
3. 可见光图像 RGB
4. 红外图像 IR
5. 文本描述线索（规划扩展）
边缘感知与预处理层
1. YOLO 行人检测
2. 视频逐帧解析
3. 行人框裁剪
4. 红外图像增强
5. 候选框质量过滤
核心识别模型层
A. TransReID 可见光重识别模块
Vision Transformer / Overlap Patch / JPM / SIE
B. DEEN 跨模态重识别模块
RGB-IR 双分支 / DEE / MFA / 正交约束
C. CLIP 图文检索启动（设计扩展）
文本编码 / 语义映射 / 以文搜图
高维身份特征
特征库与匹配层
1. 目标特征提取
2. 多图特征集合
3. L2 归一化
4. target_features.npz
5. 余弦相似度匹配
视频候选行人身份特征
6. 阈值筛选
7. 目标身份判定
输出与协同层
1. 目标锁定与高亮显示
2. 相似度分数标注
3. 结果视频输出
4. 跨摄像头轨迹辅助
5. 特征文件上传 / 下载 / 同步
6. 云边协同目标模板更新
面向公共安全、智慧城市、失踪人员查找与夜间监控补盲场景

多通道隐蔽信息传输系统
Multi-channel covert information transmission system
系统概览 发送消息 接收解密 历史溯源 隐号分析 区域网络 图片发送 图像提取 节点监控 图像生成 文件收发 用户中心
FISCO BCOS NODE OK
IPFS GATEWAY ACTIVE
SECURE SESSION: ACTIVE
ID: TEACHER1 (USER)
各节点加密算法分布
admin1
teacher1
SM2国密 CP-ABE 环签名
实时在线区块链节点总数
0 0 0 0 0 0 0 4
日活跃用户 0
累计上链密文 26
安全指数 98%
防拦截成功率 100%
今日故障数 0
实时高频通信节点 Top10
admin1 26
teacher1 26
student1 1
今日数据概况
累计上链密文 26
今日活跃用户 0
拦截尝试(已规避) 0
IPFS 存证次数 53
SM2 使用率
系统模块耗时分析
IPFS存证 链上通信 SNR降噪 拦截处理 异常告警
近6个月上链量趋势
4/13 4/14 4/15 4/16 4/17 4/18 4/19
ENCRYPTION: SM2/SM4 HYBRID
LATENCY: 24MS
BLOCK HEIGHT: 10,425
© 2026 Multi-channel covert information transmission system

智稷安防：多模态校园霸凌检测与安全预警系统架构
1 数据采集层
校园监控视频
聊天文本 / 举报文本
学生语音
图片 / 人脸 / 心理画作
测试数据与上传文件
2 终端接入层
A. 微信小程序端
教师登录 首页 行为分析 心理健康评估 霸凌事件处理 视频分析结果
B. 安保端 Web 展示
监控列表 告警区域 事件趋势分析 视频回放
3 后端服务与接口层
1. Flask + OpenCV 视频流服务
/test /video_feed /analysis_result
2. HTTP / wx.request 接口
文件上传 结果轮询 事件列表获取 复判反馈提交
3. 微信云函数
mmc 登录校验 xiaoxi 订阅消息推送
4 多模态 AI 模型层
1. 视频行为识别
Uni-AdaFocus
MobileNetV2
Violence / Nonviolence 分类
2. 中文有害文本检测
ToxiCN
Chinese RoBERTa
3. 语音情绪识别
emotion2vec
4. 图像与人脸分析
Face-ViT
5 风险融合与辅助决策层
1. 多模态结果汇总
2. 风险评分
3. 疑似霸凌判断
4. 置信度与详细分析
5. 人工复核入口（人机协同复核）
风险融合与模型优化（反馈闭环）
6 事件管理与处置闭环
1. 事件记录
2. 告警推送
3. 视频回放
4. 处置记录
5. 误判反馈
6. 趋势统计与可视化
7. 教师 / 安保 / 管理人员响应
多模态检测
人机协同复核
发现—告警—处置记录闭环管理


PhishGuard AI：基于 AI 的钓鱼邮件检测系统架构图
面向中文为主、兼顾英文的多模态钓鱼邮件检测与持续学习系统
1. 用户与入口层
用户 / 安全分析师
邮件样本上传 (.eml)
API 提交
后续可扩展：IMAP / SMTP 网关
2. 前端管理控制台
Dashboard 邮件检测页 检测详情页 人工复核页 知识库管理 模型管理 报表中心 系统配置
3. 后端服务层 (FastAPI)
FastAPI 后端 API
检测 API 反馈 API 知识库 API 模型治理 API 报表 API 系统配置 API
4. 邮件解析与预处理层
邮件解析 Header 解析 正文 / HTML 解析 URL 提取 附件提取 图片提取 SPF / DKIM / DMARC 解析
5. 多模态特征工程层
文本语义特征 Header 技术特征 URL 特征 附件特征 视觉特征 (OCR) 知识库特征 规则 / IOC 特征
6. AI 推理与规则检测层
规则引擎
SPF失败 / Reply-To不一致 / 紧急措辞 / 可疑附件 / IOC命中
文本模型
MacBERT / Chinese-RoBERTa
结构化模型
LightGBM
视觉 OCR
PaddleOCR / PP-OCR
相似案例检索
BGE-M3 / sentence-transformers + FAISS
二阶增强
CLIP / SigLIP ByT5 / CANINE Qwen / DeepSeek (解释辅助)
7. 融合决策与处置层
Fusion 融合决策模型 (LightGBM)
风险输出
risk_score / risk_level / action
分级处置
allow / warn / manual_review / quarantine / quarantine_and_learn
可解释原因
reasons / evidence
8. 数据、知识库与持续学习层
PostgreSQL
MinIO
原始邮件 / 附件 / 图片 / 模型文件
Redis + Celery
异步任务 / 重训任务 / 知识库同步
知识库
IOC 库 / 规则库 / 案例库 / 培训素材库
持续学习闭环
人工复核 → 样本池 → 离线训练 → 模型评估 → 审批上线 → 回滚
模型评估与报表
Accuracy / Precision / Recall / F1 / FPR / PNR / ROC-AUC / PR-AUC / 延迟 / 对抗测试
读 / 写 数据与知识
知识库
IOC 库 / 规则库 / 案例库 / 培训素材库
前端展示
检测结果 / 处置建议 / 可解释原因
人工复核
确认结果 / 标注反馈
架构特点：多模态特征工程 + 规则/模型混合检测 + 人机协同 + 动态知识库 + 持续学习

智慧监测平台
2025/04/21 周一 21:17:04
校园管控
数据监测
视频监控
门禁管理
设备管理
事件处置
事件处理
事件记录
事件归档
统计分析
决策分析
趋势分析
统计报表
数据挖掘
预测分析
17.13%
同比增长率
9.49%
平均响应时间(秒)
事件配置
规则配置
事件类型
报警级别
联动配置
系统管理
用户管理
权限管理
系统设置
日志管理

shGuard AI
入口安全检测工作台
当前 FPR 1.9%
待复核 4
态势总览
邮件检测
检测详情
人工复核
IOC 知识
规则库
案例库
模型评估
报表中心
系统配置
fusion-v1.0.0
未同步
态势总览
fusion-v1.0.0
FPR 1.9%
等待同步
系统配置已保存
刷新
邮件入口风险态势
模型 fusion-v1.0.0 正在处理检测、复核与知识库沉淀链路。
FPR 1.9%
P95 320 ms
今日检测 148
已完成邮件解析与风险判定
钓鱼邮件 38
进入隔离或复核范围高风险样本
高风险邮件 21
由中短段与模型判高风险集合
当前 FPR 1.9%
误报率控制目标 < 5%
误报 / 漏报 3 / 1
人机协同纠偏结果
响应延迟 191 ms
P95 320 ms
最近风险趋势
检测总量、钓鱼样本与可疑样本的 7 日变化。
总量 钓鱼 可疑
05-15 05-16 05-17 05-18 05-19 05-20 05-21
风险等级分布
用于观察当前样本结构和处置压力。
严重 16.22%
低风险 35.13%
高风险 23.9%
中风险 27.7%